# Does competence predict performance? : standardized patients as a means to investigate the relationship between competence and performance of general practitioners

# DOES COMPETENCE PREDICT PERFORMANCE?

## Standardized patients as a means to investigate the relationship between competence and performance of general practitioners

# DOES COMPETENCE PREDICT PERFORMANCE?

## Standardized patients as a means to investigate the relationship between competence and performance of general practitioners

**PROEFSCHRIFT**

ter verkrijging van de graad van doctor
aan de Rijksuniversiteit Limburg te Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. M.J. Cohen,
volgens het besluit van het College van Dekanen,
in het openbaar te verdedigen op
donderdag, 31 oktober 1991 om 14.00 uur

door

Jan-Joost Rethans

Promotoren: Prof. Dr. F. Sturmans
Prof. Dr. M.J. Drop
Co-promotor: Dr. C.P.M. van der Vleuten

Beoordelingscommissie:
Prof. Dr. C.P.A. van Boven, voorzitter
Prof. Dr. J.A. Knottnerus
Dr. J.C.M. Metz
Prof. Dr. H.G. Schmidt
Prof. Dr. Th. B. Voorn

Publication of this thesis was supported by a grant from E. Merck Nederland B.V.

..... Ik bestudeerde de kaart (touwtje langs de wegen, vermenigvuldigen met de schaal), ik reed het parkoers met mijn auto, met de auto van een vriend, ik monteerde een kilometerteller, maar al die metingen gaven een andere uitkomst; het te meten object had alleen maar de gebrekkigheid van mijn meetmethodes aan de kaak gesteld. ....

.....I studied the map (stretching a piece of string along the roads and multiplying according to the scale), I drove the course in my own car and in a friend's; I fitted a mileometer. But all these measurements gave different results; the object to be measured had only served to expose the imperfections of my measuring methods. ....

Tim Krabbé in his book "De Renner"
Erven Thomas Rap, Vijverhof, Baarn, 1980, p. 39

## Preface

This book originated as spin-off from one of my hobbies, namely long-distance running.
While running I get ideas, which usually tend to come at random or to relate to the persons and things I pass.
Sometimes, however, it is possible to concentrate on one particular issue.
During one of my training runs I was thinking about methods to open up the general practitioner's "black box", his/her consulting room.
One of the brainwaves which occurred to me was to get information about consultations by using simulated patients, who were to be sent to doctors' practices without the doctors knowing.
At first I laughed at the idea and rejected it as funny, not serious, and certainly not feasible.
However, the thought persisted and I decided to accept the challenge to start working with simulated (or standardized) patients in real practice.
It is therefore with the greatest pleasure that I am able to present this book.
It is my hope that those who read it will enjoy it and will pick up at least one general idea: conducting research can be great fun!

Contents

# Chapter 1: Introduction

This book contains the results, in the form of several papers, of two research projects conducted consecutively by the Department of General Practice at the University of Limburg in Maastricht.

The projects had two main aims. The first of these was to investigate whether it was possible for a set of standardized (or simulated) patients to visit general practitioners during normal surgery hours without being detected. The second aim was to investigate 1) the difference and 2) the relationship between competence and performance of general practitioners.

In this chapter both topics, the method of using standardized patients and the competence and performance of physicians, will be briefly introduced.

## The methodology of using standardized patients

The immediate instigation for the use of standardized patients in real practice in the projects described here was a sentence in a medical paper.[1] This referred to the hypothesis that there might be a discrepancy between what doctors actually do in their practice and what they say they do in an interview or in a questionnaire.

To measure what goes on in real practice requires instruments, which need to be valid and consistent. One of the best instruments in this regard is the use of standardized (or simulated) patients.* Numerous studies have been published dealing with the use of this method for educational and licensing purposes in medical schools. In a recent review of the use of standardized patients, evidence is given for the validity and consistency of the method.[2] Only a few studies have been published of the use of standardized patients in actual practice.[3-10]

The use of standardized patients in real practice has received much criticism in medical journals.[11-15] This criticism focussed mainly on the fact that the consent of the participating physicians was not obtained and on the lack of consistency (reliability) of the patients used.

A very strong point with regard to the use of standardized patients in real practice is that this makes it possible to compare doctors in their care for

* It was the neurologist Barrows who was the first to speak of and report on simulated patients. (Barrows HS. The programmed patient: a technique for appraising student performance in clinical neurology. J Med Educ 1964; 39: 802-5). He discussed so called "programmed patients", meaning laymen and women who had been trained to play a role as a patient in a confrontation with medical students or physicians. In 1971 Barrows changed the term "programmed patients" into "simulated patients" (Barrows HS. Simulated patients. Springfield (Ill.): Thomas, 1971). Through the years other names have been used, such as "pseudopatients" or "surrogate patients". Since 1987 there has more or less been a consensus to use the term "standardized patients" (Barrows HS. Simulated (standardized) patients and other human simulations. Chapel Hill (North Carolina): Health Sciences Consortium, 1987).

patients, with the standardized patient as the independent variable and the doctors as the dependent one.

The lack of experience with this method in actual practice was one of the challenges to be met in the projects described here. With respect to this, the objective of the studies in this book was to investigate whether the method could be used in actual practice in a valid and consistent way. In this respect, the projects can be described as innovative and risky. This method made it possible to investigate the hypothesis that there might be a discrepancy between what doctors actually do in their practice and what they say they do in an interview, questionnaire or a paper-and-pencil test (first project, chapter two).

## Competence and performance

The assessment of people's achievements is not only interesting in the area of sports, but is an even more serious business in the academic world of medicine.
Before a medical student can start a career as a medical doctor, she or he has to pass the necessary examinations. When this has been done one can speak of a competent doctor.

But does the passing of examinations tell us anything about the future performance of doctors? Intuition suggests that this should be the case but, surprisingly, no evidence exists to validate this assumption. Most of the studies intended to investigate the relationship between pre-examination and post-examination performance show a lack of sound methodological criteria.[16] For example, they either do not define the setting in which the participating physicians were assessed during the studies or they compare results collected with different instruments in the two settings.

A study was therefore needed which would investigate the relationship between competence and performance in a methodologically sound way.

Unfortunately, the literature on 'competence' makes use of an extensive vocabulary. Ability, behaviour, clinical judgment, clinical reasoning, habitual performance, problem solving, clinical competence, the combination of knowledge, skills and attitudes, naturalistic behaviour, medical competence, adequate performance, efficient performance, interpretive skills, optimal behaviour, criteria setting and clinical decision making are terms used (generally without definition) to describe this domain.

In this book the following definitions are used: performance is "what a doctor does in his day-to-day practice" and competence is "what a doctor is capable of doing".[17,18] There were several reasons for choosing these definitions.

First, they are mutually exclusive: they do not overlap. Second, the definitions leave the readers free to choose the type of measuring instruments

with which to assess aspects of either competence or performance. That is, the two definitions apply to the settings of the doctors' professional activities and not to the domain of expertise. One might argue that competence is made up of several aspects and that it would be more appropriate to speak of 'competences', referring to competences as a series of parameters such as diagnostic and therapeutic skills, interpersonal skills, etc. In this book however we want to emphasize the concept of competence as a single entity. With the measurement of competence, or performance, both concepts can be operationalized into several aspects.

Third, both definitions appeal to the intuitive feeling that working in actual practice is different from working under examination conditions and for that reason are easy to understand.

## Standards of care

Besides the lack of methodologically sound studies with which to investigate the relationship between competence and performance, there is another reason why the studies described here were initiated. There have been a number of studies, showing that doctors in real practice perform below established standards of care.[9,10, 19-21]

An interesting phenomenon in these studies is that it appears to make no difference whether the standards are made by colleagues, experts or even by the participating doctors themselves. That is, below-standard performance seems to be a consistent finding, regardless of the origin of the criteria. This was elegantly demonstrated by Norman, who asked physicians to construct standards of care which were (and this was stressed to the physicians) meant for practical use in actual care.[9] After doing so, all the participating physicians were visited in their practice by standardized patients. The results of the study showed that the doctors as a group performed only 56% of the actions considered to be essential according to the standard.

This observation of below-standard performance by doctors has in the past led to action being taken by various professional organisations. This action was mostly directed towards mandatory attendance at continuing medical education. The underlying assumption of such directives is that poor performance is a reflection of inadequate knowledge and/or skills, which are assumed to be capable of being remedied by additional instruction.

Studies regarding the effects of postgraduate education on the actual practice behaviour of doctors often produce conflicting evidence.[22] It has been shown, for instance, that additional postgraduate education does not seem to change the practice behaviour of doctors.[23] A recent study concerning postgraduate teaching of fundoscopy to general practitioners showed that there was no measurable learning effect, although the doctors were very enthusiastic.[24] In a study whose purpose was to describe the effects of training general practitioners in the application of standards, Grol showed

that the participating physicians performed 51% of necessary actions after the project in contrast with 45 % beforehand.[20] As a side effect of this project, however, doctors also performed more intermediate and superfluous actions, whereas it was expected that the incidence of such actions would be lowered after the project.

If, however, doctors do not lack competence, then the process of setting standards can also be questioned. This issue is interesting, since in recent years there has been a movement in general practice towards the designing of standards of care for practical purposes.

The designers of these standards seem to be convinced that by producing standards the quality of care in actual practice can be raised. With that in mind, the studies in this book also attempt to determine whether doctors can do better in a test situation, according to established consensus standards, than in their practical work. If doctors can in fact do better, it would be interesting to see whether they perform in line with standards or not.

## Structure of this book

As mentioned before two studies with standardized patients have been conducted. The first study was conducted in 1986 when 48 general practitioners were consulted by standardized patients (chapters two and three). In 1989 a second study was conducted during which 39 general practitioners were consulted by standardized patients (chapters seven and eight).

Since the book is divided in chapters, which are based on papers, it is possible for the reader to read only the chapter or chapters which seem to be of interest. If one is interested enough to read all the papers, one will inevitably find some overlap between some of them.

Two of the papers (chapter 3 and 4) were written especially for Dutch readers and were published in Dutch medical journals. However, chapter 3 contains a extended summary in English and for chapter 4 (with an abstract in English) several alternative papers exist in English.[2, 25].

**Chapter 2** (Simulated patients in general practice: a different look at the consultation) contains the results of our first study with standardized patients. The purpose of this paper is to compare the actual performance of doctors ("what the doctor does") with answers with regard to a written case of a standardized patient ("what the doctor says he would do").
**Chapter 3** (De dokter onderzocht /The doctor examinated. Simulated patients presenting micturition complaints in office encounters) contains results from the same study as in chapter 2, now focussing on the feasibility of sending standardized patients into general practitioners' offices.

**Chapter 4** (Simulatiepatienten in onderwijs en praktijk, een literatuur overzicht/ Simulated patients in medical education and medical practice) is a review paper, focussing on establishing the purposes for which standardized patients have been used and examining the reliability and validity of the use of standardized patients.
**Chapter 5** (Competence and performance: two different concepts in the assessment of quality of medical care). This is a review paper concerning performance and competence. Its main purpose was to look for definitions of the words "performance" and "competence", and to see if there was any empirical evidence for a relationship between them, using the definitions of performance ("what a doctor does in his day-to-day practice") and of competence ("what a doctor is capable of doing").
**Chapter 6** (A method for introducing standardized patients into general practice consultations) describes in detail the results of the study of whether it is possible for a set of standardized patients to consult general practitioners repeatedly without being detected, in a health care system in which general practitioners have fixed lists of patients.
**Chapter 7** (Assessment of the performance of general practitioners by the use of standardized patients) This chapter attempts to determine whether the actual performance of general practitioners, as assessed by standardized patients, meets predetermined standards for actual practice.
**Chapter 8** (Does competence of physicians predict their performance?) A direct comparison between an examination setting and actual practice using undetected standardized patients.) contains the results of a study whose purpose was to determine whether competence (defined as "what a doctor is capable of doing") and performance (defined as "what a doctor does in his day-today practice" ) are related.
**Chapter 9** (Conclusions and recommendations for research) rehearses and discusses the main conclusions of the work with standardized patients presented in this book and ends with some advice for further research with standardized patients.

## Appendices

1. Detection form for physicians to register detected standardized patients
2. Instructions for the competence section of the study
3. The standards of care used

## References


1. Tielens VCL, Mokkink H, Mesker P. Het medisch diagnostisch handelen van de huisarts. Huisarts Wet 1985; 28: 44-8.
2. Vleuten CPM van der, Swanson DB. Assesment of clinical skills with standardized patients: state of the art. Teaching and Learning in Medicine 1990; 2: 58-76.

3. Rosenhan DL. On being sane in insane places. Science 1973; 179: 250-79.
4. Owen A, Winkler R. General practitioners and psychological problems: an evaluation using pseudopatients. Med J Austr 1974; 2:393-8.
5. Winkler RC. Research into mental health practice using pseudopatients. Med J Aust; 2: 399-403.
6. Amidi S, Solter S, Rashidian B, Zokajan A, Rasmjolan F. Antibiotic use and abuse among physicians in private practice in Shiran, Iran. Med Care 1975; 13: 341-5.
7. Burri A, McCaughan K, Barrows H. The feasibility of using the simulated patient as a means to evaluate clinical competence in a community. Proc Annu Conf Res Med Educ 1976; 15: 295.
8. Renaud M, Beauchemin J, Lalonde C, Poirier H, Berthiaume. Practice settings and prescribing profiles: the simulation of tension headache to general practitioners working in different practice settings in Montreal area. Am Publ Health 1980; 70: 1068-73.
9. Norman GR, Neufeld VR, Walsh A, Woodward CA, McConvey GA. Measuring physicians' performances by using simulated patients. J Med Educ 1985; 60: 925-34.
10. McClure CL, Gall EP, Meredith KE, Annel Gooden M, Boyer JT. Assessing clinical judgment with standardized patients. J Fam Pract 1985; 20: 457-64.
11. Anonymus. Show Biz (Editorial). JAMA 1973; 225:59.
12. Anonymus. Pseudopatients (Editorial). Br Med J 1974; 268: 729.
13. McConaghy. Pseudopatients and evaluation of medical care. Med J Aust 1974; 2: 385.
14. Anonymus. Simulated patients (Editorial). Br Med J 1974; 286: 399-400.
15. Anonymus. Pseudopatients and pseudoresearch (Editorial). JAMA 1975; 232: 59.
16. Rethans JJ, Leeuwen v Y, Drop R, Sturmans F, Vleuten vd C. Performance and competence: two different constructs in the assessment of quality of medical care. Fam Pract 1990; 7: 168-74.
17. Senior JR. Towards the measurement of competence of medicine. Philadelphia: National Board of Medical Examiners 1976.
18. Lloyd JS. Definitions of competence in specialties of medicine. Chicago: American Board of Medical specialties 1979.
19. Norman GR, Tugwell P, Feightner JW. A comparison of resident performance on real and simulated patients. J Med Educ 1982; 57: 708-15.
20. Grol R. Kwaliteitsbewaking in de huisartsgeneeskunde. Krips Repro Meppel, 1986. Thesis. (Summary in english).
21. Lomas J, Anderson GM, Domnick-Pierre K, et al. Do practice guidelines guide practice? N Engl J Med 1989; 321:1306-11.
22. Haynes R, Davis D, Mc Kibbon A, Tugwell P. A critical appraisal of the efficacy of continuing medical education. JAMA 1984; 251: 61-64.
23. Sibley JC, Sackett DL, Neufeld V, Gerrard B, Rudnick KV, Fraser W. A randomized trial of continuing medical education. NEJM 1982; 306: 511-5.
24. Reenders K, Hiemstra RJ, Bender W. Teaching fundoscopy to GP's in an OSCE-setting: failures and successes. In: Bender W et al (eds). Teaching and assessing clinical competence. Groningen 1990, Boekwerk Publications: 374-8.
25. Barrows HS, Simulated (Standardized) patients and other human simulations. Health Sciences Consortium. Chapel Hill, North Carolina, 1987.

## Chapter 2: Simulated patients in general practice: a different look at the consultation

Reprinted from the **British Medical Journal**, 28 March 1987, 294, 809- 812

The paper has also been published as:
1. Rethans JJE, van Boven CPA. Simulatiepatienten in de huisartspraktijk: het consult anders bekeken. Ned Tijdsch Geneeskunde 1988; 132: 1162-6.
2. Rethans JJE, van Boven CPA. Simulatiepatienten in de huisartspraktijk: het consult anders bekeken. Consult van de huisarts 1988; 12: 34-37.

# Simulated patients in general practice: a different look at the consultation

J J E RETHANS, C P A VAN BOVEN

**Abstract**

**To develop a better empirical basis for developing quality assessment in general practice three simulated patients made appointments with 48 general practitioners during actual surgery hours and collected facts about their performance. The simulated patients were indistinguishable from real patients and presented a standardised story of a symptomatic urinary tract infection. Two months later the same general practitioners received a written simulation about a patient who had the same urinary tract infection and were asked how they would handle this in real practice. Both results were scored against an existing consensus standard. The overall score for both methods did not show any substantial differences. A more differentiated analysis, however, showed that general practitioners performed significantly better with simulated patients. It also showed that general practitioners answering the written simulation performed significantly more unnecessary and superfluous actions.**

**The results of this study show that the use of simulated patients seems to show the efficient performance of general practitioners in practice.**

Department of General Practice, University of Limburg, Postbus 616, 6200 MD Maastricht, The Netherlands

J J E RETHANS, MD, general practitioner

Department of Medical Microbiology, University of Limburg, Maastricht, The Netherlands

C P A VAN BOVEN, PHD, professor

Correspondence to: Dr Rethans.

**Introduction**

In general practice as well as in hospitals efforts are being made t develop "ideal" standards and equivalent, logical, branchin flow charts for clinical performance. These developments aim a improving the performance of individual tasks, to promote a mor uniform performance by doctors, and to provide standards for th quality of care in solving clinical problems.[1,2] The use of suc standards has important consequences—for example, a positive o negative evaluation of an individual doctor's performance and reduced cost of health care for the community.

Most methods for developing standards are based on consensu models, in which a group of general practitioners discusses th complaint or medical problem in question and decides on a standard procedure for dealing with it. Studies of the use of such standards, however, show that doctors seem to do less well than such standards deem to be desirable.[3,4] General practitioners may not be acting as efficiently as the designers of the standard would like, but the standard may also be too academic in that it does not allow for the circumstances actually occurring in general practice. It is therefore desirable to establish a standard of performance for general practitioners based on the conditions of real life.

Methods of assessing actual performance may be either indirect, by sending a questionnaire to general practitioners to answer questions about problems with managing patients, or direct, by actually observing doctors while they are consulting. Comparisons have shown a discrepancy between doctors' clinical performance as assessed by indirect methods and what they actually do when they are observed in their practice.[3] Observed doctors omit actions that were expected from them from the results of the indirect methods. These studies, however, compared different groups of doctors dealing with different medical problems.

Though there is a substantial experience with the use of simulated patients for educational purposes,[5-8] and the simulated patient method has been described as the best method to assess the

management of patients by doctors,[9] few studies have used simulated patients to research problems in actual practice.[9-12] This study aimed at comparing the performances of a group of general practitioners when dealing with a particular medical problem by an indirect method of assessment, using written details about a simulated patient, and a direct method, using simulated patients in practice. We hypothesised that the general practitioners would do less in a consultation with the simulated patient than they would have indicated in their answers to the written problem.

The study therefore had the following purposes: to investigate the possibility of using trained simulated patients to collect data about the reality of the consultation; to measure the difference between the performances of general practitioners when dealing with the simulated patient (what the doctor does) and the actions that the same general practitioners said that they would take when dealing with a written problem about a similar patient (what the doctor says he would do); and to measure the agreement between an already existing standard based on a consensus model and the actual performance of general practitioners as shown by data gathered using simulated patients.

## Subjects and methods

The medical problem in this study concerned a 30 year old woman who complained about painful micturition for one and a half days. This problem was constructed as a standardised role for three simulated patients, who were paid to participate in this project and had no medical background. The problem was also written down in the form of a "simulation of initial medical problem solving."

A *During your surgery Mrs Cox appears. You don't know her: she is with her husband and two children on holiday. To you she looks about 30 years old. She complains of painful micturition for the past day and a half.*

What would you like to ask her about her history?

B *During your surgery hours Mrs Smith appears. She is 30 years old, and you have not seen her for a long time. She lives with her husband and two children in their own house in a new part of the town. She tells you that she has been well for the past few years. She now complains of painful micturition for the past day and a half. You question her and suspect that she has an uncomplicated urinary tract infection.*

1 If you usually perform a physical examination what exactly would you do?
2 If you usually perform laboratory tests what exactly would you do?

*After your completed history and examination you conclude that she has indeed got an uncomplicated urinary tract infection.*

3 What would be your further management (in terms of what you would prescribe and for how long, what you would tell her, and how you would arrange a follow up)?

Written simulation of initial medical problem solving used in study.

The written form, meant to be used to assess medical competence, was a simplification of a modified essay question (E de Graaf, unpublished observations). The figure shows the format used in this study. The written format made it necessary to mention the suspected urinary tract infection before the doctor continued with the questions on management (during a real consultation the doctor receives answers from the patient). For this reason part A and part B were presented separately to the doctors (while reading part A they were unable to see part B).

The facts collected about history, results of physical and laboratory examination, instructions given to the patient, treatment, and follow up were scored according to an existing standard on urinary tract infection developed at the University of Nijmegen.[13] The standard is divided into obligatory actions (considered to be necessary) and intermediate actions (not essential but not harmful either). All other actions are considered to be superfluous (see tables I and II for the items listed in the standard used).

The simulated patients were trained in their role as a patient and in reporting reliable and valid facts about the consultation. To check the reliability among the simulated patients we assessed three test consultations. The patients completed a checklist of items (based on the standard) immediately after such a consultation, and the same consultation was also scored by three independent doctors. For each consultation the scores obtained by the three independent doctors were considered to be the gold standard for that particular consultation. Subsequently, the score of the simulated patient was compared with this gold standard to assess their reliability, which was 1·0 ($\kappa$=1·0), 0·89 ($\kappa$=0·78), and 0·89 ($\kappa$=0·78), respectively. To assess the consistency of individual patients each patient was retested after six weeks with the same consultation (recorded on videotape). Their consistencies were 0·96 ($\kappa$=0·92), 0·89 ($\kappa$=0·78), and 0·85 ($\kappa$=0·70), respectively. During the study the simulated patients were themselves responsible for making appointments with the doctors. They presented themselves as Mrs Cox (figure).

Most Dutch general practice units have their own small laboratory, where medical secretaries perform simple tests. Simulated patients are obviously unable to undergo such tests, so we asked the secretaries of the participating doctors what kind of laboratory investigations their doctor would order them to do to diagnose a urinary tract infection. The reliability of the answers received was assessed by asking the secretaries the same question one week later. The agreement between their first and second answers was 99·25%.

TABLE I—*Number (%) of general practitioners (n=25) performing items listed in the standard procedure for urinary tract infection during consultation with simulated patient and in answers to written simulation*

| | Item No | Obligatory(O), intermediate (F), or superfluous (S) action | Performed: During consultation | Performed: In written answer |
|---|---|---|---|---|
| History: | | | | |
| Chief complaint | 1 | O | 25 (100) | 25 (100) |
| Frequency | 2 | O | 24 (96) | 18 (72)* |
| Onset of complaint | 3 | O | 13 (52) | 11 (44) |
| Ask about general wellbeing | 4 | O | 14 (56) | 11 (44) |
| Ask if this is first time | 5 | F | 22 (88) | 22 (88) |
| Ask for history of urogenital tract problems | 6 | F | 4 (16) | 10 (40) |
| Vaginal discharge | 7 | F | 5 (20) | 11 (44) |
| Superfluous actions | 8 | S | 5 (20) | 12 (48) |
| Physical examination: | | | | |
| Measure blood pressure | 9 | F | 1 (4) | 1 (4) |
| Feel flank for pain | 10 | F | 3 (12) | 9 (36) |
| Superfluous actions | 11 | S | 5 (20) | 13 (52)* |
| Instructions to patient: | | | | |
| Explain diagnosis | 12 | O | 23 (92) | 6 (24)** |
| Explain prognosis | 13 | O | 12 (48) | 2 (8)* |
| Advise voiding after intercourse | 14 | F | 1 (4) | 2 (8) |
| Advise drinking more fluids | 15 | F | 11 (44) | 9 (36) |
| Discuss micturition habits | 16 | F | 2 (8) | 5 (20) |
| Ask for fresh midstream specimen | 17 | O | 1 (4) | |
| Superfluous actions | 18 | S | 3 (12) | 2 (8) |
| Treatment: | | | | |
| Antibiotics | 19 | O | 25 (100) | 23 (92) |
| Superfluous actions | 20 | S | | |
| Follow up: | | | | |
| Standard | 21 | O | 20 (8) | 19 (76) |
| After treatment | 22 | O | 18 (72) | 16 (64) |
| If complaint persists despite drugs return earlier | 23 | O | 1 (4) | 3 (12) |
| Superfluous actions | 24 | S | 1 (4) | |

*p<0·05, **p<0·0005.

TABLE II—*Number (%) of medical secretaries (n=25) performing items listed for laboratory examination in standard procedure for urinary tract infection, and number (%) of supervising general practitioners (n=25) indicating in answers to written simulation that these items are performed*

| | Item No | Obligatory(O), intermediate (F), or superfluous (S) action | Performed: By secretary | Performed: In written answer |
|---|---|---|---|---|
| Clean midstream urine specimen | 1 | O | 1 (4) | 2 (8) |
| Microscopical urine examination | 2 | O | 21 (84) | 20 (80) |
| Urine concentration of: | | | | |
| Glucose | 3 | O | 21 (84) | 13 (52) |
| Nitrite | 4 | F | 9 (36) | 10 (40) |
| Albumin | 5 | F | 23 (92) | 14 (56)* |
| Dipslide or urine culture | 6 | F | 1 (4) | |
| Superfluous actions | 7 | S | 1 (4) | |

*p<0·05.

TABLE III—*Number and range of actions listed in standard procedure for urinary tract infections performed by general practitioners (n=25) during consultation with simulated patient and in answers to written simulation (excluding laboratory examination). Each item of standard procedure has value of one point*

| Actions | Mean (SD) No performed | Range | % Of standard |
|---|---|---|---|
| Total: | | | |
| Simulated patient | 9·88 (3·44) | 4-19 | |
| Written simulation | 10·04 (3·37) | 4-18 | |
| Obligatory: | | | |
| Simulated patient | 7·04 (1·54) ** | 4-9 | 64 |
| Written simulation | 5·24 (1·80) ** | 2-8 | 47 |
| Intermediate: | | | |
| Simulated patient | 1·96 (1·24) * | 0-6 | 24 |
| Written simulation | 2·76 (1·36) * | 0-6 | 34 |
| Superfluous: | | | |
| Simulated patient | 0·88 (1·74) * | 0-8 | |
| Written simulation | 2·04 (2·16) * | 0-8 | |

*p<0·05, **p<0·005, Student's *t* test for paired design.

All 378 general practitioners working in the same county as our university received information about this study four and a half months before the actual consultations with the simulated patients took place. They were told about the possibility of a visit by a simulated patient, who would collect facts about the consultation, during the next 12 months. The content of the case was specifically not mentioned. If a doctor did not want to participate in the study he or she had to withdraw actively; otherwise we considered him or her to be a possible participant.

Because one of us (JJER) was teaching at a school for medical secretaries during this study we made the criterion for including a general practitioner in the study the fact that he or she employed a secretary still connected with the school. This opened up the possibility of collecting information directly from the secretaries about the laboratory part of the consultation without bothering their supervising general practitioners and without disclosing the actual design of the study.

The participating doctors were visited by the simulated patients over a period of four and a half weeks. Two months after the visits all doctors were sent the written simulation and were requested to return it to us. The written problems that were returned were scored by a doctor according to the urinary tract infection standard. A select sample of the scores was also scored again independently by another doctor. Their coefficient of agreement was 0·94 ($\alpha$=0·89).

The two tailed McNemar test for paired data was used to compare the proportion of general practitioners performing an action during a consultation with a simulated patient and the proportion of general practitioners who explicitly said that they would carry out the same action in reply to the written simulation (significant at $p<0·05$). The two tailed Student's *t* test for paired data was used to compare the total number of actions scored by the general practitioners with the two methods (significant at $p<0·05$).

## Results

Of the 378 doctors receiving notification of the study, 30 actively withdrew, among them three of the original 69 doctors who fitted the inclusion criterion. Of the remaining 66 doctors, 14 were excluded because of having connections with the department of general practice of our university or because their practice was based too far away. Of the remaining 52 doctors chosen for visits, four could not be visited: one simulated patient knew the doctor that she was supposed to visit, one patient was unable to get an appointment with her doctor, and two doctors could not be visited within the time allowed for this study. Thus at the end of the study 48 general practitioners had been visited. Of the 378 doctors receiving written simulation, 219 (57%) returned it, including 27 (56%) of the 48 doctors who were visited. Two of these 27 wrote spontaneously to say that they had detected the simulated patient; these two were excluded from the statistical analysis.

Table I shows, for each item of the urinary tract infection standard, the proportion of general practitioners who actually performed the action as reported by the simulated patient compared with the proportion indicating by the written method that they would perform the action. Table II compares separately the comments made by the doctors and their secretaries about laboratory tests. We do not present the results of all 48 doctors who were visited because 21 did not return the written simulation, and we cannot therefore make any comparison. The proportion of doctors per item of these 48 general practitioners, however, does not deviate substantially from that of the 25 doctors reported on here.

Table III shows the mean number of actions performed as reported by the simulated patients compared with the number performed as assessed by the written method. It shows also the number performed as a percentage of the number recommended by the standard.

## Discussion

This study shows that simulated patients can produce reliable and valid data about the performance of doctors in actual practice. Only two of the 48 general practitioners who were visited reported afterwards that they detected the simulated patient. We have no reason to suspect that others also did, though we admit that we cannot be sure. Some doctors, on receiving the written simulation, may have recognised the visit of the simulated patient because they were able to recall the features of the case. To prevent the doctors from recalling the consultation with the simulated patient we introduced Mrs Smith in the written method. It could be argued that the section on management in the written simulation no longer applies to the visit of the simulated patient Mrs Cox. This is true, but there is no strict medical reason for the management of Mrs Smith and Mrs Cox to differ in the items recorded here.

Table I shows clearly that in taking the patient's history, giving the patient instructions, prescribing treatment, and following up the doctors performed more obligatory actions during the visit by the simulated patient than indicated by their corresponding intentions in the written simulation (except for item 23). The total score for both methods, however (table III), shows that more acts were intended to be performed than were performed in real life, but the difference was minimal and not significant. A more differentiated analysis, however, shows that for the obligatory acts doctors do more in reality than they say that they intend to do as measured by the written simulation, a finding that is at variance with our initial hypothesis. For the intermediate and superfluous actions, however, the findings were just the opposite.

These results are surprising. Most surprising is the finding that there seems to be a clear distinction between obligatory actions on the one hand and intermediate and superfluous actions on the other. We think that perhaps during the consultation with the simulated patient, under the pressure of time, the doctors deal efficiently with a routine problem. During the written simulation, however, they tend to want to show their knowledge, exhibiting an examination technique and forgetting about normal routine procedures. The actions and intentions for laboratory examinations show fewer differences, but this is the only category in which doctors are not being compared with themselves. Such results give reason to believe that using simulated patients shows a pattern of efficient performance (for the most essential issues) by general practitioners.

We are well aware that content specificity could affect the results.[14] More research is therefore necessary to investigate whether the simulated patient method also shows this special pattern of efficient performance when more doctors are dealing with other medical problems. Practical problems, such as devising different ways to enter practices and deciding on which medical problems can be used in a simulation, still have to be resolved.

This study shows that the use of simulated patients in practice results in more (and probably more relevant) information about the performance of general practitioners than other existing methods. This is shown most clearly by the category "instructions to patients." The answers obtained by the indirect method would lead to the conclusion that doctors forget to tell their patients about the diagnosis, while the direct method shows that doctors are giving excellent instructions.

During consultation with a simulated patient the participating doctors performed only 64% and 24% of the obligatory and intermediate actions, respectively (table III). We have no reason to suspect the participating doctors of having low standards of care. We therefore think that these results indicate that the consensus standard does not take into account sufficiently the reality of practice for general practitioners. The standard should have a better empirical basis. The simulated patient method could be the method of providing this basis. Standards should almost certainly be

evolved in a practical setting even if developed initially in an abstract manner by discussion groups. In any case, the measurement of performance against standards should be carried out practically and not by an abstract questionnaire.

The results of this study encourage the use of the simulated patient method as a basis for establishing the actual performance patterns of general practitioners. The full documentation of these patients in daily practice still seems to be a great challenge.

We thank D L Crombie, C de Geus, J Hendrix, P Hobus, A Knottnerus, C Phaff, M Verwijnen, and W van Zutphen for their help during the study.

## References

1 Irvine D. Quality of care in general practice: our outstanding problem. *J R Coll Gen Pract* 1983;33:521-3.
2 Wulff HR. *Probabilities and value judgements in clinical decision-making. Evaluation of efficacy of medical action.* Amsterdam: North-Holland Publishing, 1979.
3 Tielens VCL, Mokkink H, Mesker P. Het medisch diagnostisch handelen van de huisarts. *Huisarts en Wetenschap* 1985;28:44-8.
4 Norman GR, Tugwell P, Feighter JW. A comparison of resident performance on real and simulated patients. *J Med Educ* 1982;57:708-15.
5 Norman GR, Feighter JW. A comparison of behaviour on simulated patients and patient management problems. *Med Educ* 1981;15:26-32.
6 Lamont CT, Hennen BKE. The use of simulated patients in a certification examination in family medicine. *J Med Educ* 1972;47:789-92.
7 Barrows HS. *Simulated patients.* Springfield, Illinois: Charles C Thomas, 1971.
8 Goran MJ, Williamson JW, Gonella JS. The validity of patient management problems. *J Med Educ* 1973;48:171-7.
9 Norman GR, Neufield VR, Walsh A, Woodward CA, McConvey GA. Measuring physicians' performances by using simulated patients. *J Med Educ* 1985;60:925-34.
10 Owen A, Winkler R. General practitioners and psychosocial problems: an evaluation using pseudopatients. *Med J Aust* 1974;2:393-8.
11 Burri A, McCaughan K, Barrows HS. The feasibility of using the simulated patient as a means to evaluate clinical competence of practising physicians in a community. *Proc Annu Conf Res Med Educ* 1976;15:295.
12 McClure CL, Gall EP, Meredith KE, Annel Gooden M, Boyer JT. Assessing clinical judgment with standardized patients. *J Fam Pract* 1985;20:457-64.
13 Grol R, Mesker P, eds. *Huisarts en toetsing.* Nijmegen: Nijmeegs Universitair Huisartsen Institut, 1984.
14 Elstein AS, Shulman LS, Sprafka SA. *Medical problem solving: an analysis of clinical reasoning.* Cambridge, Massachusetts: Harvard University Press, 1978.

*(Accepted 23 January 1987)*

## Chapter 3: De dokter onderzocht: simulatiepatienten met mictieklachten op het spreekuur van de huisarts (The doctor examined. Simulated patients presenting micturition complaints in office encounters)

Reprinted from **Huisarts en Wetenschap**, 1988, 31, 3-6

## Summary

A simulated patient is a person who has rehearsed a role as a patient, which role he then plays in an encounter with a doctor. Initially simulated patients were used only during medical vocational training in (self) evaluation of trainees; later, they were also used in real practice situations. [7-9] In an effort to establish the feasibility and effectiveness of this method for Dutch general practices, a study was carried out of the performance of general practitioners confronted with the complaint 'painful micturition'. The casuistics chosen concerned a woman aged about 30 who had for the first time developed a symptomatic urinary tract infection during a holiday and presented to an unfamiliar general practitioner with the complaint 'painful micturition'.

For reporting the data on the encounter the protocol 'Micturition complaints' of the Nijmegen University department of General Practice was used.[16] All items relevant to the casuistics chosen were culled from this protocol, and this list of items was used as measuring instrument for the simulated encounter.

The study was limited to 69 general practitioners who had a trainee-practice aide completing a period of practical training in their practice. Via the school for practice aides, information was obtained on the laboratory tests used by these general practitioners when confronted with micturition complaints.

From a pool of some 100 simulated patients three women aged about 30 were selected who were considered able to give a detailed account of a simulated encounter afterwards. These women were carefully trained to play their role and to report reliably on an encounter.

The simulated patients were given samples of urine to take to the encounter. It was ascertained that on the day of the encounter all had similar, unmistakably infected urine samples with them, divided into two portions. If required, the woman could give the general practitioner one portion; she could give the other portion if she was asked to produce urine 'on the spot'. A seventh, identical urine sample was tested at the university microbiological laboratory for nitrite, glucose, albumin, pH, leucocytes and sediment (more or less at the time of the women's encounter in general practice). An urinary culture was started as well.

Of the initial population of 69 general practitioners, 23 dropped out for various reasons. Of the 46 remaining general practitioners, 78 percent posed only three anamnestic questions and only one asked for a washed midstream sample. Other striking findings were:

- 35 percent of the general practitioners used nitrofurantoin as therapeutic agent of choise;
- only 19 percent of the general practitioners prescribed an antimicrobial course of less than seven days;

- of the general practitioners who made an appointment for a follow-up, 92 percent asked the patient to report back 5-10 days after the start of the medication;
- 40 percent of the 'mandatory' items were omitted;
- active general practitioners were actice in all categories; those who did little, did little in all categories.

According to the guidelines of the protocol used, an item was characterized as mandatory if it involved a truly necessary activity; its omission was regarded as negligence. In this study this would imply that the general practitioners were negligent with regard to 40 percent of the mandatory items. The study population consisted of general practitioners who met the criteria formulated by the schools for practice aides. Consequently there is no reason to assume that especially these general practitioners would provide poor medical care; it seems more probable that the Nijmegen protocol, despite exhaustive validation, accounts insufficiently for the reality of an encounter; in other words: the standard set is too high. The findings raise questions, both regarding the way in which standards are set and regarding the extended training of general practitioners.

# De dokter onderzocht

## Simulatiepatiënten met mictieklachten op het spreekuur van de huisarts

JAN-JOOST RETHANS EN CEES VAN BOVEN

**Drie getrainde simulatiepatiënten bezochten 48 huisartsen tijdens hun spreekuur in een geblindeerde setting. Zij presenteerden zich als patiënten met een symptomatische urineweginfectie en registreerden wat de huisartsen tijdens het consult vroegen en deden. Doel van dit onderzoek was om te zien of de simulatiepatiënt-methode haalbaar en betrouwbaar is in de Nederlandse huisartspraktijk en om het werkelijke beleid van huisartsen bij een patiënt met een urineweginfectie vast te stellen. De resultaten werden gescoord met behulp van het protocol 'Mictieklachten' van het Nijmeegs Universitair Huisartsen Instituut. De simulatiepatiënten bleken zeer betrouwbaar een consult weer te kunnen geven en slechts twee van de 48 huisartsen ontmaskerden de simulatiepatiënt. De andere 46 huisartsen verrichtten slechts 60 procent van de handelingen die volgens het protocol 'obligaat' waren.**

## Inleiding

Protocollaire geneeskunde en intercollegiale toetsing zijn actuele ontwikkelingen in de huisartsgeneeskunde die vragen om methoden die de werkelijkheid van het spreekuur zichtbaar maken.[1,2] In de visie van het Nederlands Huisartsen Genootschap dienen de in het kader van de protocollering te ontwikkelen normen aan te sluiten bij de huidige gang van zaken in de praktijk en tevens wetenschappelijk verantwoord te zijn.[3] Ook *De Haan en Hollenbeek Brouwer* komen tot de conclusie dat theoretische normen pas werkelijk betrouwbaar zullen zijn, als zij in de praktijk worden geijkt: 'De normen van het medisch handelen dienen te worden ontleend aan het veld van praktizerende huisartsen'.[4] Dit houdt in dat het noodzakelijk is om betrouwbare informatie over het handelen in de praktijksituatie te verkrijgen.

Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.
J.J.E. Rethans, huisarts, vakgroep Huisartsgeneeskunde; Prof. dr. C.P.A. van Boven, hoogleraar, vakgroep Medische Microbiologie.
Correspondentie: J.J.E. Rethans.

Voor het verkrijgen van gegevens over het handelen van huisartsen tijdens het dagelijks spreekuur zijn verschillende methoden ontwikkeld:
- enquête;
- analyse van het kaartsysteem (*chart audit*);
- papieren-patiëntenproblemen;
- contact met een simulatiepatiënt in een laboratoriumsituatie, eventueel op video vastgelegd;
- observatie van het spreekuur door een observator, eventueel gecombineerd met een audio-opname;
- contact met een simulatiepatiënt tijdens het spreekuur, waarbij de simulatiepatiënt gegevens over het consult verzamelt.

In welke mate deze methoden de werkelijkheid van het spreekuurcontact weergeven, is overigens vaak niet duidelijk. Zo werd een discrepantie gevonden tussen de gegevens die via observatie waren verkregen, en de rapportage door de huisartsen zelf.[5]

Een simulatiepatiënt – de naam en methodiek zijn bedacht door *Barrows*[6] – is iemand die een rol als patiënt heeft ingestudeerd en deze rol vervolgens speelt in een contact met een arts. Soms worden echte patiënten gebruikt, die dan hun eigen ziekte in een gestandaardiseerde rol presenteren. Aanvankelijk werden simulatiepatiënten alleen tijdens de medische opleiding gebruikt ter (zelf)evaluatie van studenten, later ook in echte praktijksituaties.[7-9]

Het toepassen van voor de arts onbekende en niet herkende simulatiepatiënten in de dagelijkse praktijk met het doel informatie over zijn handelen te verkrijgen heeft scherpe kritiek ondervonden, met name omdat het onethisch zou zijn artsen zonder hun medeweten te observeren.[10,11] Een probleem is ook de betrouwbaarheid van de rapportage van de simulatiepatiënt over een consult.[12,13] Recent is echter aangetoond dat systematische oefening van simulatiepatiënten tot een aanvaardbaar niveau van betrouwbaarheid leidt.[14,15]

Om de toepasbaarheid en bruikbaarheid van deze methode voor de Nederlandse huisartspraktijk vast te stellen, werd een onderzoek verricht naar het handelen van huisartsen bij de klacht 'pijn bij het plassen'.

## Methoden

### Mictieklachten

De gekozen casuïstiek had betrekking op een circa 30-jarige vrouw die voor de eerste keer een symptomatische urineweginfecties had en zich tijdens haar vakantie meldde bij een haar vreemde huisarts meldt de klacht 'pijn bij het plassen'. Bij het instuderen van de rol werd, behalve aan het correct naar voren brengen van de klacht, veel aandacht besteed aan het verhaal om de klacht heen (vakantie, verzekeringsvorm, beroep, etc).

De gegevens die door de simulatiepatiënt over het consult zouden worden gerapporteerd, moesten uiteraard eenduidig worden vastgelegd. Hiervoor is gebruik gemaakt van het protocol 'Mictieklachten' van het Nijmeegs Universitair Huisartsen Instituut (NUHI).[16] Voor dit protocol is gekozen, omdat de ontstaanswijze – literatuurstudie, consensus-methode en uittesten in de praktijk – een garantie leek voor een goede aansluiting op de praktijk. Uit het protocol werden alle items gedestilleerd die op de gekozen casuïstiek betrekking hadden (*tabel 1*); deze itemlijst is gebruikt als meetinstrument voor het simulatiecontact.

### Huisartsen

Alle 378 huisartsen in de regio van de Rijksuniversiteit Limburg kregen ongeveer vier maanden voor het onderzoek een brief waarin uitleg over de vraagstelling werd gegeven en waarin werd meegedeeld dat te zijner tijd simulatiepatiënten een aantal praktijken zouden bezoeken en zouden registreren wat de bezochte huisartsen deden tijdens het consult. Nadrukkelijk werd niet vermeld om welke klacht het ging. Huisartsen die bezwaar tegen het onderzoek hadden, konden zich afmelden; bij geen bericht werd aangenomen dat men toestemming verleende.

Omdat het beleid bij mictieklachten wordt kenmerkt door een groot aantal handelingen dat zich in het laboratorium afspeelt, was het noodzakelijk om ook over dit onderdeel goede informatie te verkrijgen. Daarom is het onderzoek in eerste aanleg beperkt tot de 69 huisartsen die een stagiaire van een doktersassistentenopleiding in hun praktijk hadden. Via de school (en met toestemming van de opleiding) werd

informatie ingewonnen over de laboratoriumdiagnostiek bij mictieklachten van de betreffende huisartsen.

Huisartsen die moeilijk te bereiken waren, en huisartsen die op de hoogte zouden kunnen zijn van de te presenteren casuïstiek, werden uitgesloten.

**Simulatiepatiënten**

De Rijksuniversiteit Limburg beschikt via haar afdeling Skillslab over een *pool* van circa 100 simulatiepatiënten die gebruikt worden ten behoeve van het medisch onderwijs. Uit deze groep werden drie vrouwen van rond de 30 jaar gekozen, die in staat werden geacht om een simulatiecontact na afloop goed weer te geven, en bovendien gedurende zes weken vier ochtenden per week beschikbaar waren. Deze drie vrouwen werden vervolgens op het Skillslab getraind in het spelen van hun rol en het betrouwbaar weergeven van een consult.

Voor de evaluatie hiervan werden tijdens de trainingen drie meetpunten ingesteld, met tussenpozen van drie weken. Tijdens de metingen vond op het Skillslab een video-opname plaats van een spreekuurcontact van een huisarts met alle drie de simulatiepatiënten. Deze video-opnamen werden onafhankelijk van elkaar beoordeeld door drie artsen aan de hand van de genoemde itemlijst. Aldus werd een gouden standaard voor elk individueel consult vastgesteld. Anders gezegd, de antwoorden die de beoordelaars op de scoringslijst invulden, golden als de 100 procentscore van dat consult. De behaalde scores van deze interbetrouwbaarheidsmetingen waren respectievelijk 1.0 (kappa = 1.0), 0.89 (kappa = 0.78) en 0.89 (kappa = 0.78).

Naast deze interbetrouwbaarheidsmeting werd ook de intrabetrouwbaarheid van de simulatiepatiënten gemeten door ieder van hen hun eigen derde meting, die op video was opgenomen, na vier weken opnieuw te laten scoren. De patiënten ontvingen extra training op de vragen die zij onjuist invulden. De hierbij behaalde scores bedroegen 0.96 (kappa = 0.92), 0.89 (kappa = 0.78) en 0.85 (kappa = 0.75).

**Urine**

In hun rol als patiënt hadden de simulatiepatiënten urinemonsters bij zich. Er werd voor gezorgd dat alle vrouwen per onderzoeksdag dezelfde, duidelijk geïnfecteerde urine bij zich hadden en wel verdeeld over twee porties. Eén portie kon de vrouw desgevraagd overhandigen aan de arts en een tweede portie kon zij overhandigen als haar werd gevraagd om ter plekke urine te produceren. Zij goot dan op de WC de tweede portie over in het aangereikte bekkentje.

Een zevende, identiek monster van de urine werd omstreeks hetzelfde tijdstip dat de vrouwen op de praktijk waren op het microbiologisch laboratorium van de universiteit bekeken op nitriet, glucose, albumen, pH, leukocyten en sediment. Tevens werd een uricult ingezet. Aldus werd per dag een gouden urinestandaard vastgesteld.

De betrouwbaarheid van de verstrekte gegevens door de doktersassistentes werd gemeten door hen de informatie over het laboratorium tweemaal te vragen, waarbij de assistentes de eerste keer niet wisten dat dit nogmaals gevraagd zou worden. De overeenkomst tussen hun gegevens van de eerste en tweede keer vragen was 99.25 procent.

**Scores**

Naargelang het aantal verrichte handelingen hebben we de artsen scores toegekend voor hun consulten. Per verricht item werd 1 punt toegekend. Aan de hand van het aantal verrichte handelingen konden we nagaan hoe de groep huisartsen scoorde op het gebruikte protocol, zowel voor het totaal als voor de categorieën obligaat en facultatief. Omdat de categorie overbodige handelingen in principe oneindig groot is, kon hier geen percentage over worden berekend.

Om te bezien of het mogelijk was de betrokken huisartsen te beoordelen op hun werkwijze, hebben we ook geanalyseerd in hoeverre er samenhang is tus-

**Tabel 1** *De geselecteerde items uit het Nijmeegse protocol 'Mictieklachten'[16] en het percentage huisartsen (n=46) dat de verschillende items verrichtte.*

| | | |
|---|---|---|
| **Anamnese** | | |
| 1. Aard van de klacht | O | 100 |
| 2. Frequentie van de mictie | O | 89 |
| 3. Hoelang bestaat de klacht | O | 46 |
| 4. Algemeen welbevinden | O | 50 |
| 5. Recidieven nagaan | F | 78 |
| 6. Voorgeschiedenis t.a.v. tractus urogenitalis nagaan | F | 9 |
| 7. Vraagt naar fluor vaginalis/pruritus vulvae | F | 17 |
| 8. Overbodige handelingen anamnese | Ov | 20 |
| **Lichamelijk onderzoek** | | |
| 9. Tensiecontrole | F | 4 |
| 10. Slagpijn/nierloge-onderzoek | F | 11 |
| 11. Overbodige handelingen onderzoek | Ov | 13 |
| **Eigen laboratorium** | | |
| 12. Gewassen middenplas, onderzoek max. 2 uur na lozing | O | 2 |
| 13. Sediment | O | 78 |
| 14. Urine glucose | O | 83 |
| 15. Urine nitriet | F | 39 |
| 16. Dipslide of urinekweek | F | 2 |
| 17. Urine albumen | F | 91 |
| 18. Overbodige handelingen laboratorium | Ov | 2 |
| **Voorlichting** | | |
| 19. Uitleg diagnose | O | 89 |
| 20. Prognose bespreken | O | 46 |
| 21. Advies goed uitplassen na coitus | F | 2 |
| 22. Veel drinken adviseren | F | 46 |
| 23. Ingaan op mictiegewoonten: hygiëne | F | 7 |
| 24. Verse gewassen urine vragen | O | 2 |
| 25. Overbodige handelingen voorlichting | Ov | 15 |
| **Medicatie** | | |
| 26. Antibiotica medicatie | O | 100 |
| 27. Overbodige handelingen medicatie | Ov | 0 |
| **Terugbestellen** | | |
| 28. Controle afspreken | O | 78 |
| 29. Controle urine-onderzoek na afloop van de kuur | O | 72 |
| 30. Aangeven dat de patiënt bij aanhouden van de klacht (ondanks medicatie) eerder terug moet komen | O | 7 |
| 31. Overbodige handelingen controle | Ov | 2 |

O= obligaat; Ov= overbodig; F= facultatief.

sen respectievelijk het aantal obligate, facultatieve en overbodige handelingen en het totaal verrichte aantal handelingen.

## Resultaten

Van de aanvankelijke populatie van 69 huisartsen zijn er 23 afgevallen:
- 3 huisartsen meldden zich na de aankondigingsbrief af;
- 4 huisartsen werden geschrapt in verband met mogelijke bekendheid met de te presenteren klachten;
- 10 huisartsen werden niet bezocht in verband met de afstand;
- 1 huisarts weigerde de patiënt te ontvangen;
- 3 huisartsen werden niet bezocht in verband met tijdgebrek.
- 2 van de 48 bezochte huisartsen meldden achteraf dat ze er bijna zeker van waren dat ze de simulatiepatiënt hadden ontdekt.

De resultaten van de consulten van de resterende 46 huisartsen zijn vermeld in *tabel 1*. Hieruit blijkt onder meer dat 78 procent van de huisartsen slechts drie anamnestische vragen stelde, en dat slechts één huisarts om gewassen middenplas-urine vroeg.

Uit *tabel 2* blijkt dat 40 procent van de 'obligate' items niet is verricht.

In de *tabellen 3-5* zijn enkele items van het protocol gedetailleerder uitgewerkt:
- nitrofurantoine was voor 35 procent van de huisartsen een eerste-keus preparaat *(tabel 3)*;
- slechts 19 procent van de huisartsen schreef een antimicrobiële kuur voor van minder dan zeven dagen *(tabel 4)*;
- van de huisartsen die een nacontrole afspraken, maakte 92 procent die afspraak voor 5 tot 10 dagen na het begin van de medicatie *(tabel 5)*.

Uit *tabel 6* blijkt tenslotte dat artsen die in totaal veel handelingen verrichtten, in alle categorieën veel deden, en dat artsen die weinig deden, in alles weinig deden.

**Tabel 2** *Kwantitatieve gegevens betreffende de door de 46 huisartsen verrichte handelingen.*

| Handelingen | Gemiddelde | Spreiding | % van de standaard | p.25-p.75 |
|---|---|---|---|---|
| Totaal | 13,2 | 7-23 | | 10,7-15,2 |
| Obligaat | 8,41 | 4-12 | 60 | 7,0-10,0 |
| Facultatief | 4,04 | 1-9 | 36 | 3,0- 5,0 |
| Overbodig | 0,76 | 0-8 | | 0,0- 1,0 |

**Tabel 3** *De voorgeschreven medicatie en het percentage huisartsen (n=46) dat deze middelen voorschreef.*

| Medicatie | Percentage |
|---|---|
| Nitrofurantoine | 35 |
| Cotrimoxazol | 24 |
| Sulfamethizol | 15 |
| Trimetoprim | 9 |
| Pipemidinezuur | 9 |
| Noroxin | 7 |
| Amoxicilline | 2 |

**Tabel 4** *De duur van de voorgeschreven kuren en het percentage huisartsen (n=46) dat deze kuren voorschreef.*

| Aantal dagen | Percentage |
|---|---|
| 4 | 2 |
| 5 | 15 |
| 6 | 2 |
| 7 | 43 |
| 8 | 2 |
| 10 | 35 |

**Tabel 5** *Het afgesproken aantal dagen tussen consult en nacontrole en het percentage huisartsen (n=35) dat deze afspraken maakte.*

| Aantal dagen | Percentage |
|---|---|
| 5 | 6 |
| 6 | 3 |
| 7 | 34 |
| 8 | 9 |
| 9 | 3 |
| 10 | 37 |
| 14 | 6 |
| 21 | 3 |

**Tabel 6** *Samenhang (Pearson correlatiecoëfficiënt) tussen het totaal aantal door de 46 huisartsen verrichte handelingen en het aantal obligate, facultatieve en overbodige handelingen.*

| Handelingen | Totaal | P |
|---|---|---|
| Obligaat | .74 | 0.000 |
| Facultatief | .77 | 0.000 |
| Overbodig | .72 | 0.000 |

## Beschouwing

Niet eerder is op deze manier informatie verkregen over het feitelijk handelen van huisartsen. In die zin zijn deze resultaten dan ook uniek.

Volgens de richtlijnen van het gebruikte protocol is een handeling als obligaat gekenmerkt als het om een echt noodzakelijke handeling gaat; het weglaten ervan wordt beschouwd als nalatig handelen. In dit onderzoek zou dat betekenen dat de artsen nalatig waren bij 40 procent van de obligate items.

De onderzochte groep bestaat uit huisartsen die voldoen aan de criteria die opleidingen voor doktersassistentes stellen. Wij hebben dan ook geen reden om aan te nemen dat speciaal deze huisartsen slechte medische zorg zouden verlenen. Het lijkt ons waarschijnlijker dat het Nijmeegse protocol, ondanks de uitvoerige wijze van validering, toch te weinig rekening houdt met de werkelijkheid van een consult, dat wil zeggen dat de norm te hoog is gesteld. In de literatuur over de diagnostiek van urineweginfecties, zoals gepubliceerd in *Huisarts en Wetenschap* in de laatste 15 jaar, komt bijvoorbeeld keer op keer de gewassen middenplas-urine, onderzocht met de dipslide-methode als enige juiste methode uit de bus. In ons onderzoek heeft slechts één van de 46 huisartsen hiervan gebruik gemaakt. Dit roept vraagtekens op ten aanzien van de manier waarop normen worden gesteld, en ten aanzien van de nascholing die huisartsen ontvangen.

Via getrainde simulatiepatiënten feitelijke gegevens over het handelen van huisartsen verzamelen, is om verschillende redenen aantrekkelijk en zinvol.

Allereerst zal het moeilijk zijn een methode van dataverzameling in de praktijk te vinden die meer valide is. Juist het feit dat de huisarts in kwestie niet weet dat er tijdens een consult gegevens over zijn handelen worden verzameld, maakt dat hij de kans krijgt om zijn echte handelen te laten zien. In tweede fase van dit onderzoek bijvoorbeeld werd aan de bezochte huisartsen gevraagd wat ze in de dagelijkse praktijk zouden voorschrijven bij een ongecompliceerde urineweginfectie. In de schriftelijke antwoorden werd door 35 procent van de huisartsen sulfamethisol als eerste keus werd aangegeven; 26 procent koos voor cotrimoxazol en 17 procent voor nitrofurantoine.[18] *Boerema* vond zelfs via een enquête onder 5000 huisartsen dat ruim 50 procent van de Nederlandse huisartsen zou kiezen

voor een sulfapreparaat.[18] *Tabel 3* laat echter zien dat de samenstelling van de 'top 3' in de praktijk weleens heel anders zou kunnen zijn. Ons inziens geven de resultaten met de simulatiepatiënt-methode aan dat op zijn minst voorzichtigheid is geboden met het trekken van conclusies uit gegevens die schriftelijk bij huisartsen zijn verzameld.

*Baselier* pleit voor een kortdurende therapie van 48 uur als behandeling van urineweginfecties.[19] In de praktijk wordt dit volgens onze gegevens nauwelijks gedaan.

Een ander voordeel van het gebruik van simulatiepatiënten op het spreekuur is dat op deze manier reeds ontwikkelde protocollen voor het huisartsgeneeskundig handelen getoetst kunnen worden op hun haalbaarheid en dat andersom, zeker in het kader van de meerjarige beroepsopleiding, ook huisartsen getoetst kunnen worden op hun handelen. Door het achteraf aan huisartsen vragen waarom men wel of niet een handeling verrichtte, kan tevens inzicht ontstaan in de rationale van het dagelijks handelen.

Bij aanvang van dit onderzoek hebben de simulatiepatiënten onder grote psychische druk gestaan, vooral door de gevolgde manier van toestemming vragen aan de huisartsen. De simulatiepatiënten voelden zich ook ongemakkelijk toen zij merkten dat de artsen slecht 'scoorden' op de itemlijst, terwijl zij het gevoel hadden adequaat te worden behandeld. In de loop van het onderzoek werden de patiënten steeds meer ontspannen en tenslotte waren zij zonder meer bereid bij een volgend onderzoek weer mee te doen.

Wij realiseren ons terdege dat nog zeer veel problemen bij het invoeren van deze simulatiepatiënt-methode moeten worden overwonnen. Met name de praktische uitvoerbaarheid en de ethische problematiek vragen nog om denkwerk en discussie. Toch menen wij dat deze methode, mits verantwoord en goed gebruikt, een goed inzicht kan verschaffen in de werkelijkheid van de spreekkamer van de huisarts.

**Dankbetuiging**

Wij willen hierbij onze dank uitspreken aan C. de Geus, J. Hendrix, P. Hobus, V. Kaiser, A. Knottnerus, Y. van Leeuwen, C. Phaff en W. van Zutphen voor hun bijdrage aan deze studie.

[1] Van den Dool CWA. Pro protocol. Huisarts Wet 1984; 27(suppl 8): 37-9.

[2] Wigersma L. Doelstellingen en beperkingen van protocollen in de huisartsgeneeskunde. Huisarts Wet 1986; 29: 300-5.

[3] Van de Lisdonk E. Huisarts en kwaliteit. Huisarts Wet 1981; 24: 250-1.

[4] De Haan M, Hollenbeek Brouwer H. Normen voor het dokteren van de huisarts. Med Contact 1982; 37: 1650-3.

[5] Tielens VCL, Mokkink H, Mesker P. Het medisch diagnostisch handelen van de huisarts. Huisarts Wet 1985; 28: 44-8.

[6] Burri A, McCaughan K, Barrows H. The feasibility of using the simulated patient as a means to evaluate clinical competence of practicing physicians in a community. Rime-conference, 6, p. 295.

[7] Barrows HS. Simulated patients. Springfield (Ill.): Thomas, 1971.

[8] Owen A, Winkler R. General practitioners and psychosocial problems; an evaluation using pseudopatients. Med J Australia 1974; 2: 393-8.

[9] Rosenhan DL. On being sane in insane places. Science 1973; 179: 250-79.

[10] McConaghy N. Pseudopatients and evaluation of medical practice. Med J Aust 1974; 2: 385.

[11] Anonymous. Pseudopatients and pseudoresearch [Editorial]. JAMA 1975; 232: 59-60.

[12] Lamont CT, Hennen BKE. The use of simulated patients in a certification examination in family medicine. J Med Educ 972; 72: 789-95.

[13] Norman GR, Tugwell P, Feighter JW. A comparison of resident performance on real and simulated patients. J Med Educ 1982; 57: 708-15.

[14] McClure CL, Gall EP, Meredith KE, Annel Gooden M, Boyer JT. Assessing clinical judgment with standardized patients. J Fam Pract 1985; 20: 457-64.

[15] Norman GR, Neufeld VR, Walsh A, Woodward CA, McConvey GA. Measuring physicians' performances by using simulated patients. J Med Educ 1985; 60: 925-34.

[16] Grol R, Mesker P, red. Huisarts en toetsing. Nijmegen: Nijmeegs Universitair Huisartsen Instituut, 1984.

[17] Rethans JJE, Van Boven CPA. Simulated patients in general practice: a different look at the consultation. Br Med J 1987; 294: 9-12.

[18] Boerema JBJ. Behandeling van urineweginfecties. TGO 1985; 10: 731-6.

[19] Baselier PJAM. Acute bacteriële urineweginfecties in de huisartspraktijk [Dissertatie]. Nijmegen: Katholieke Universiteit Nijmegen, 1983.

## Chapter 4: Simulatiepatienten in onderwijs en praktijk, een literatuur overzicht (Simulated patients in medical education and medical practice. A literature survey)

Reprinted from **Huisarts en Wetenschap**, 1989, 32, 366-369

## Abstract

Since its introduction by Barrows in 1964, the simulated patient method has been used in medical education and for research and assessment purposes in medical practice. Inter-reliability in the use of simulated patients for assessment purposes has not yet been satisfactorily investigated. This yields also to what specifically can be concluded from these assessments. The method is however well documented for educational purposes and for registration of actual procedures in medical practice and its use in these simulations is justified. The simulated patient method offers the possibility of more closely investigating inter-doctor variation in general practice.

# Simulatiepatiënten in onderwijs en praktijk

## Een literatuuroverzicht

JAN-JOOST RETHANS
RIET DROP
FERD STURMANS
YVONNE VAN LEEUWEN

**Sinds de introductie van de simulatiepatiënt in 1964 door Barrows is deze methode gebruikt in het medisch onderwijs en voor onderzoek en toetsing in de dagelijkse praktijksituatie. De interbetrouwbaarheid bij het gebruik van simulatiepatiënten bij toetsing is nog onvoldoende onderzocht. Verder zal bij toetsing duidelijker moeten worden wát er geconcludeerd kan worden uit gegevens die verkregen worden via simulatiepatiënten. Voor educatieve doeleinden en voor registratie van de feitelijke gang van zaken in de medische praktijk is de methode echter goed gedocumenteerd en is het gebruik ervan gerechtvaardigd. Voor de huisartspraktijk biedt de simulatiepatiënt-methode onder meer mogelijkheden om het begrip variatie nader te onderzoeken.**

**Rethans JJ, Drop R, Sturmans F, Van Leeuwen Y. Simulatiepatiënten in onderwijs en praktijk. Een literatuuronderzoek. Huisarts Wet 1989; 32(10): 366-9.**

Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.

J.J. Rethans, huisarts, vakgroep Huisartsgeneeskunde; prof. dr. R. Drop, vakgroep Medische Sociologie; prof. dr. F. Sturmans, vakgroep Epidemiologie; Y. van Leeuwen, huisarts, vakgroep Huisartsgeneeskunde.

Correspondentie: J.J. Rethans.

### Inleiding

In de afgelopen jaren is in ons land geëxperimenteerd met het consulteren door simulatiepatiënten van praktizerende huisartsen met het doel informatie over het consult te verkrijgen.[1,2] De ervaring met deze nieuwe vorm van praktijkonderzoek is nog beperkt, maar gelet op het streven naar meer praktijkgericht onderzoek, is te verwachten dat in de toekomst simulatiepatiënten vaker gebruikt zullen worden om feitelijke praktijksituaties te onderzoeken.

In dit artikel geven wij een overzicht van de verschillende aspecten van het gebruik van simulatiepatiënten. In het bijzonder komen twee vragen aan de orde:

- Voor welke doeleinden zijn simulatiepatiënten tot nu toe binnen de geneeskunde gebruikt?
- Hoe betrouwbaar en hoe valide is het gebruik van simulatiepatiënten?

Voor de beantwoording van deze vragen is gebruik gemaakt van ons bestand aan publikaties over simulatiepatiënten. De basis daarvan wordt gevormd door het bestand, zoals dat in november 1986 bestond aan de McMaster Universiteit (Hamilton, Canada), destijds toonaangevend op het gebied van onderzoek van en met simulatiepatiënten. Na 1986 hebben wij dit bestand systematisch aangevuld met de relevante literatuur.

### Barrows

Bijna 25 jaar geleden publiceerde de neuroloog *Barrows* als eerste over simulatiepatiënten.[3] Hij besprak een methode voor het testen van medisch studenten: de 'programmed patient'-methode. *Barrows* omschreef een 'programmed patient' als 'een leek die een rol als patiënt heeft ingestudeerd en deze rol vervolgens speelt in een contact met een medisch student of een arts'.

*Barrows* gaf twee redenen om met 'programmed patients' te werken bij het evalueren van medisch studenten. Ten eerste zijn deze 'patiënten' gestandaardiseerd, in tegenstelling tot echte patiënten: iedere student ziet dezelfde patiënt en de klinische vaardigheden van de studenten zijn daardoor beter met elkaar te vergelijken. Ten tweede hoeft bij een ontmoeting tussen een student en een getrainde 'patiënt' geen observator aanwezig te zijn, zodat de student zich niet direct geobserveerd hoeft te voelen.

In 1971 verving *Barrows* het woord 'programmed' door 'simulated'[4] en in de loop der jaren zijn er nog verscheidene andere benamingen geweest, zoals 'pseudopatients', 'surrogate patients' en 'standardized patients'. In alle gevallen gaat het echter om de oorspronkelijke 'geprogrammeerde patiënt' van *Barrows*.

Sinds 1964 zijn simulatiepatiënten gebruikt voor onderwijsdoeleinden in de medische studie en voor onderzoek naar het praktisch medisch handelen van artsen.

### Onderwijs

Het heeft lang geduurd voordat de methodiek van *Barrows* navolging vond; men dacht in het begin dat simulatiepatiënten slechts een zeer beperkt aantal rollen zouden kunnen spelen.[5] Tot het eind van de zeventiger jaren zijn dan ook bijna geen publikaties van belang te vinden over onderwijskundige ervaringen met simulatiepatiënten. Sindsdien is het aantal publikaties echter sterk toegenomen.

Binnen het onderwijs heeft de simulatiepatiënt-methode twee functies gekregen: een educatieve en een evaluatieve functie.

*Godkins et al.* zijn begonnen met het gebruik van simulatiepatiënten bij het aanleren van gynaecologisch onderzoek.[6] Aanvankelijk had de patiënt daarbij alleen een passieve rol, maar door *Kretzschmar* is die rol uitgebreid tot een actieve participatie, waarbij de patiënt feedback gaf over zowel de medisch-technische als de affectieve aspecten van het gynaecologisch onderzoek.[7] Recent heeft *Van Luijsen* verslag gedaan van ervaringen op dit terrein in Groningen.[8]

Baanbrekend werk is vooral verricht door *Stillman et al.*[9,10] Zij leidden simulatiepatiënten op tot 'patient-instructors'. De term 'patient-instructor' houdt in dat een simulatiepatiënt getraind wordt in een rol waarbij meestal lichamelijk onderzoek moet worden verricht. De patiënt onder-

gaat het onderzoek en geeft vervolgens feedback over de prestatie van de student. *Stillman et al.* toonden aan dat het mogelijk is de 'patient-instructor' zo te trainen dat deze een betrouwbaar oordeel over studenten kan geven.[10] Ook *Johnson et al.* en *Gerber et al.* maken gebruik van 'patient-instructors'.[11 12]

Over het aanleren van sociale vaardigheden is gerapporteerd door *Helfer et al.* en *Owen and Underwood*.[13 14] Zij stellen dat contacten met simulatiepatiënten unieke leermomenten opleveren op het gebied van de sociale vaardigheden. *Maguire et al.* komen tot een vergelijkbare conclusie met betrekking tot het het aanleren van een goede anamnesetechniek.[15]

De evaluatieve functie heeft aanvankelijk minder aandacht gekregen. In 1972 rapporteerden *Lamont and Hennen* over het gebruik van simulatiepatiënten bij examens in de huisartsgeneeskunde[16] en pas in 1979 werd opnieuw gepubliceerd over simulatiepatiënten bij toetsing.[9] De meeste publikaties op dit gebied zijn van zeer recente aard.[17-23] Inmiddels wordt echter op tal van medische faculteiten gebruik gemaakt van simulatiepatiënten bij het toetsen van studenten.

## Praktijk

De meest tot de verbeelding sprekende toepassing van simulatiepatiënten is het gebruik in de dagelijkse praktijk van artsen in een geblindeerde situatie. Tot nu toe zijn er negen onderzoeken op dit terrein gepubliceerd.

*Rosenhan* stuurde acht 'gezonde' individuen naar twaalf psychiatrische klinieken om te kijken of de psychiaters konden differentiëren tussen deze 'patiënten' en 'echte' psychiatrische patiënten.[24] De simulatiepatiënten simuleerden psychiatrisch gedrag tot en met het moment van opname; direct daarna staakten ze dit gedrag. De opnameduur varieerde van 7 tot 52 dagen. Alle simulatiepatiënten werden ontslagen onder de diagnose 'schizofrenie in remissie'.

In de periode 1974-1976 werden vier studies verricht, waarbij simulatiepatiënten huisartspraktijken bezochten en het feitelijke praktijkgedrag registreerden.[25-28] Deze onderzoeken leidden in de medische wereld tot een storm van protest, waarbij de voornaamste kritiek gericht was op het feit dat tevoren geen toestemming was gevraagd aan betreffende artsen.[29-33] Overigens wordt in geen van deze studies een uitspraak gedaan over de betrouwbaarheid van de rapportage.

Pas na 1980 werd de lijn van praktijkonderzoek opnieuw opgepakt. *Renaud et al.* stuurden simulatiepatiënten met hoofdpijn naar een aantal overheidspraktijken en privé-praktijen, *McClure et al.* onderzochten de behandeling van gewrichtsklachten door huisartsen met behulp van simulatiepatiënten, *Norman et al.* gebruikten verschillende klachten, zoals nekpijn en buikpijn, en *Rethans en Van Boven* gebruikten mictieproblemen.[34-37 1 2]

## Betrouwbaarheid

Of simulatiepatiënten betrouwbaar moeten zijn (en in welke mate), hangt af van het doel waarvoor ze worden gebruikt.

We kunnen vier aspecten aan de betrouwbaarheid onderscheiden:
- speelt een simulatiepatiënt dezelfde rol bij herhaling op dezelfde manier (intra-betrouwbaarheid);
- spelen verschillende simulatiepatiënten dezelfde rol op dezelfde manier (inter-betrouwbaarheid);
- zijn simulatiepatiënten bij gebruik voor toetsingsdoeleinden in staat om een betrouwbaar oordeel over studenten te geven;
- zijn simulatiepatiënten bij praktijkonderzoek, in staat om betrouwbaar te rapporteren wat er in de praktijk gebeurt?

Bij gebruik van simulatiepatiënten voor educatieve doeleinden speelt betrouwbaarheid praktisch geen rol, omdat in die situaties studenten en simulatiepatiënten vrij moeten kunnen 'experimenteren'. Anders ligt het bij het gebruik van simulatiepatiënten voor toetsings- en praktijkdoeleinden; daar speelt betrouwbaarheid juist een belangrijke rol.

Studies die expliciet tot doel hadden om de betrouwbaarheid van simulatiepatiënten te onderzoeken, zijn zeer schaars.[38]

In 1987 rapporteerden *Vu et al.* over een onderzoek waarin wordt nagegaan of drie verschillende personen consistent dezelfde rol konden spelen en of één persoon tijdens verschillende sessies op de dag consistent zijn rol bleef volhouden.[39] Beide betrouwbaarheidsprecentages lagen boven de 90 procent. *Dawson-Saunders et al.* vonden echter verschillen tussen simulatiepatiënten die dezelfde rol speelden. Zij benadrukken dat een goede training van simulatiepatiënten essentieel is.[40] Verder is in onderwijskundige toetsingssituaties aangetoond dat het oordeel van simulatiepatiënten over studenten even betrouwbaar kan zijn als het oordeel van observatoren.[41] Tot slot zijn er wat betreft het rapporteren over de medische praktijk vier studies waarin de betrouwbaarheid van de rapportages door simulatiepatiënten is onderzocht (gemiddeld een kappa van .85).[1 35-37]

## Validiteit

In educatieve situaties is het van belang dat simulatiepatiënten zoveel mogelijk op echte patiënten lijken. Voor het valide registreren van het gedrag van artsen is dit zelfs de belangrijkste voorwaarde.

*Sanson-Fisher and Poole* onderzochten of studenten consulten met afwisselend echte patiënten en simulatiepatiënten empathisch verschillend ondergingen.[42] *Norman et al.* onderzochten of artsen in een gedragslaboratorium simulatiepatiënten konden onderscheiden van echte patiënten.[43] In beide studies konden de proefpersonen de simulatiepatiënten niet van de echte onderscheiden. In overeenstemming hiermee is de bevinding van *Woodward et al.* en *Rethans and Van Boven* dat hun simulatiepatiënten slechts zelden ontmaskerd werden (tijdens respectievelijk 13 en 4 procent van de consulten).[37 1]

Bij registratie van praktijkgedrag is het tevens van belang dat de registratie en terugrapportage weergeven wat de arts feitelijk doet en niet wordt gekleurd door de sympathie of antipathie van de simulatiepatiënt. De trainingen die simulatiepatiënten krijgen, zijn erop gericht te voorkomen dat persoonlijke voorkeuren een rol spelen, maar tot op heden is naar dit aspect nog geen onderzoek verricht.

Ook bij het gebruik van simulatiepatiënten voor toetsing geldt dat simulatiepatiënten zoveel mogelijk op echte patiënten moeten lijken. Er zijn inmiddels vele studies verricht waarin verslag wordt gedaan van succesvol gebruik van simulatiepatiënten als toetsingsinstrument.[17-23 41 44-47] Overigens bestaat nog steeds onenigheid over de vraag wát er nu precies met simulatiepatiënten gemeten wordt. Dit is onder andere een gevolg van het moeilijk meetbare begrip 'medische competentie'. In dit opzicht onderscheidt de simulatiepatiënt-methode zich niet van andere meetinstrumenten als geschreven simulaties, *multiple choice*-toetsen en kennistoetsen. Indicatief voor deze situatie is dat *Neufeld and Norman* in hun boek 'Assessing clinical competence' klinische competentie een 'battered child' noemen.[48]

*Stillman and Gillers* vermelden in een overzicht van evaluaties van klinische competentie weliswaar dat simulatiepatiënten in de loop der jaren gebruikt zijn om alle aspecten van klinische competentie te meten – met name interpersoonlijk vaardigheden en interview- en probleemoplossingsvaardigheden – maar maken toch niet duidelijk hoe valide die metingen waren.[49] Tot nog toe zijn er geen onderzoeken geweest die een eenduidige conclusie over dit onderwerp toelaten.

## Beschouwing

Wat kan nu worden geconcludeerd ten aanzien van de betrouwbaarheid, de validiteit en het toekomstig gebruik van de simulatiepatiënt-methode?

Voor educatieve doeleinden zijn simulatiepatiënten een uniek leermiddel; met name de studies van *Stillman et al.* bevestigen dit keer op keer. Het gebruik van simulatiepatiënten op onderwijskundig terrein zal daarom hoogstwaarschijnlijk toenemen. Daarbij zullen ook de mogelijkheden worden uitgebreid, met name door verbetering van simulaties op lichamelijk gebied. Het gebruik van simulatiepatiënten voor onderwijskundige toetsing en praktijkdoeleinden (registratie en toetsing) vereist echter ook nader onderzoek.

Hoewel *Stillman et al.* hebben aangetoond dat 'patient-instructors' studenten kunnen beoordelen op een manier die niet onderdoet voor het beoordelen door observatoren, is de betrouwbaarheid van simulatiepatiënten bij toetsing pas zeer recent expliciet onderzocht.[39 40] Deze onderzoeken wekken weliswaar vertrouwen, maar zijn nog te gering in aantal. Het is daarom aan te bevelen dat er meer studies naar de interbetrouwbaarheid van simulatiepatiënten zullen worden opgezet. Er zijn met name nog geen studies verricht die alle aspecten van de betrouwbaarheid tegelijkertijd onder de loep nemen.

De betrouwbaarheid van simulatiepatiënten bij het rapporteren over het handelen van artsen in de dagelijkse praktijk is beter onderbouwd. In recente studies haalden simulatiepatiënten in alle gevallen betrouwbaarheidspercentages van meer dan 85 procent.[1 35 36]

De validiteit van simulatiepatiënten voor educatieve en registratie-doeleinden is voldoende onderbouwd. Wat betreft de toetsing zal duidelijker gemaakt moeten worden wát er nu precies wordt gemeten (bijvoorbeeld kennis of sociale vaardigheden).

Tenslotte de vraag, wat simulatiepatiënten voor de huisartspraktijk kunnen betekenen. Allereerst is er geen ander 'instrument' dat zo valide kan weergeven wat zich in de spreekamer afspeelt, en dat bovendien ook andere informatie verschaft dan bijvoorbeeld via schriftelijke instrumenten kan worden verkregen. Inzicht in de variatie tussen huisartsen kan beter worden bestudeerd, doordat een simulatiepatiënt keer op keer dezelfde patiënt blijft. De huisarts wordt zo de echte afhankelijke variabele. Daarnaast kunnen simulatiepatiënten in de praktijk een uitstekend onderwijsmedium zijn bij de opleiding van assistenten in de huisartsopleiding.


### Dankbetuiging

Met dank aan drs. Cees van der Vleuten van de afdeling Onderwijsontwikkeling en -research van de Rijksuniversiteit Limburg voor zijn waardevolle suggesties.


### Literatuur


1. Rethans JJE, Van Boven CPA. Simulated patients in general practice: a different look at the consultation. Br Med J 1987; 294: 809-12.
2. Rethans JJ, Van Boven C. De dokter onderzocht: simulatiepatiënten op het spreekuur van de huisarts. Huisarts Wet 1988; 31: 3-6.
3. Barrows HS. The programmed patient: a technique for appraising student performance in clinical neurology. J Med Educ 1964; 39: 802-5.
4. Barrows HS. Simulated patients. Springield (Ill.): Thomas, 1971.
5. Barrows HS. Simulated (Standardized) patients and other human simulations. Chapel Hill (North Carolina): Health Sciences Consortium, 1987.
6. Godkins TR, Duffy D, Greenwood J, Stanhope WD. Utilization of simulated patients to teach the 'routine' pelvic examination. J Med Educ 1974; 49: 1174-8.
7. Kretzschmar RM. Evolution of the Gynecology Teaching Associate: An education specialist. Am Journal Obstet Gynecol 1978; 131: 367-73.
8. Van Lunsen HW. Wie is er bang voor het gynaecologisch onderzoek? [Dissertatie]. Groningen: Rijks Universiteit Groningen, 1986.
9. Stillman P, Ruggill J, Rutala P, Sabers D. An instructional program using patient instructors as teachers and evaluators. Proc Annu Conf Res Med Educ 1979; 18: 53-8.
10. Stillman PL, Ruggill JS, Rutala PJ, Sabers DL. Patient instructors as teachers and evaluators. J Med Educ 1980; 55: 186-93.
11. Johnson BF, Perkowski L, Lanier R. Evaluation of self-instructional training program for simulated patients. Proc Annu Conf Res Med Educ 1979; 18: 47-52.
12. Gerber W, Albanese M, Brown D, Matthes S. A study of the long term effectiviness of education provided to medical students by teaching associate simulated patients. Proc Annu Conf Res Med Educ 1983; 22: 180-5.
13. Helfer E, Black MA, Teitelbaum H. Pediatric interviewing skills using real and simulated mothers. Pediatrics 1987; 55: 397-400.
14. Owen A, Underwood P. Videotape and simulated patients. Med J Aust 1980; 1: 437-40.
15. Maguire GP, Clarke D, Jolley B. An experimental comparison of three courses in history-taking skills for medical students. Med Educ 1977; 11: 175-82.
16. Lamont CT, Hennen BKE. The use of simulated patients in a certification examination in family medicine. J Med Educ 1972; 72: 789-95.
17. Coggan PG, Knight P, Davis P. Evaluating students in family medicine using simulated patients. J Fam Pract 1980; 19: 259-65.
18. Ramsey D, Darosa D, Finch WT, Konrad R, Birtch A, Peters MJ, Folse J. Evaluation of

surgical residents using simulated patients. In: Hart IR, Harden RM, eds. Further developments in assessing clinical competence. Montreal: Heal Publications, 1987.
[19] McClure CL, Gall EP, Meredith KE, Gooden MA, Boyer JT. Family practice and internal medicine clinical judgment in a university setting. J Fam Prac 1986; 22: 443-8.
[20] Grava-Gubbins J, Khan SB, Rainsberry P. Factor analysis of simulated office oral examinations in family medicine. In: Hart IR, Harden RM, eds. Further developments in assessing clinical competence. Montreal: Heal Publications, 1987.
[21] Willams RG, Barrows HS, Vu NV, Verhulst SJ, Colliver JA, Marcy M, Steward D. Direct, standardized assessment of clinical competence. Med Educ 1987; 21: 482-9.
[22] Petrusa ER, Blackwell T, Parcel S, Saydjari C. Psychometric properties of the Objective Clinical Exam (OCE) as an instrument for final evaluation. In: Hart IR, Harden RM, eds. Further developments in assessing clinical competence. Montreal: Heal Publications, 1986.
[23] Stillman PL, et al. Assessing clinical skills of residents with standardized patients. Ann Inter Med 1986; 105: 762-771.
[24] Rosenhan DL. On being sane in insane places. Science 1973; 179: 250-79.
[25] Owen A, Winkler R. General practitioners and psychosocial problems: an evaluation using pseudopatients. Med J Aust 1974; 2: 393-8.
[26] Winkler RC. Research into mental health practice using pseudopatients. Med J Aust 1974; 2: 399-403.
[27] Amidi S, Solter S, Rashidian B, Zokajan A, Rasmjolan F. Antibiotic use and abuse among physicians in private practice in Shiraz, Iran. Med Care 1975; 13: 341-5.
[28] Burri A, McCaughan K, Barrows H. The feasibility of using the simulated patient as a means to evaluate clinical competence in a community. Proc Annu Conf Res Med Educ 1976; 15: 295.
[29] Anonymus. Show Biz [Editorial]. JAMA 1973; 225: 59.
[30] Anonymus. Pseudopatients [Editorial]. Br Med J 1974; 268: 729.
[31] McConaghy. Pseudopatients and evaluation of medical practice. Med J Aust 1974; 2: 385.
[32] Anonymus. Simulated patients [Editorial]. Br Med J 1974; 399-400.
[33] Anonymus. Pseudopatients and pseudoresearch [Editorial]. JAMA 1975; 232: 59.
[34] Renaud M, Beauchemin J, Lalonde C, Poirier H, Berthiaume. Practice settings and prescribing profiles: The simulation of tension headaches to general practioners working in different practice setting in the Montreal area. Am J Publ Health 1980; 70: 1068-73.
[35] McClure CL, Gall EP, Meredith KE, Gooden MA, Boyuer JT. Assessing clinical judgment with standardized patients. J Fam Prac 1985; 20: 457-64.
[36] Norman GR, Neufeld VR, Walsh A, Woodward CA, McConvey GA. Measuring physicians' performances by using simulated patients. J Med Educ 1985; 60: 925-34.
[37] Woodward CA, McConvey GA, Neufeld V, Norman GR, Walsh A. Measurement of physician performance by standardized patients. Med Care 1985; 23: 1019-27.
[38] Jolly B. A review of issues in live patient simulation. PLET; 1982: 99-107.
[39] Vu NV, Steward DE, Marcy M. An assessment of the consistency and accuracy of standardized patients' simulation. J Med Educ 1987; 62: 100-2.
[40] Dawson-Saunders B, Verhulst SJ, Marcy M, Steward DE. Variability in standardized patients and its effect on student performance. In: Hart IR, Harden RM, eds. Further developments in assessing clinical competence. Montreal: Heal Publications, 1987.
[41] Stillman PL, Swanson DB. Ensuring the clinical competence of medical graduates through standardized patients. Arch Intern Med 1987; 147: 1049-52.
[42] Samson-Fisher, Poole Ad. Simulated patient and the assessment of medical students' interpersonal skills. Med Educ 1980; 14: 249-53.
[43] Norman GR, Tugwell P, Feightner JW. A comparison of resident performance on real and simulated patients. J Med Educ 1982; 57: 708-15.
[44] Levine H, McGuire C. The use of role-playing to evaluate effective skills in medicine. J Med Educ 1970; 45: 700-5.
[45] Stillman PL, et al. Assessing clinical skills of residents with standardized patients. Ann Intern Med 1986; 105: 762-71.
[46] De Graaff E, Post GJ, Drop MJ. Validation of a new measure of clinical problem-solving. Med Educ 1987; 21: 213-8.
[47] Smith DH, McWhinney IR. Comparison of the diagnostic methods of family physicians and internists. J Med Educ 1975; 50: 264-70.
[48] Neufeld VR, Norman GR. Assessing clinical competence. New York: Springer, 1985.
[49] Stillman PL, Gillers MA. Evaluation of clinical competence. 1988. Submitted for publication. ■

**Abstract**

**Rethans JJ, Drop R, Sturmans F, Van Leeuwen Y. Simulated patients in medical education and medical practice. A literature survey. Huisarts Wet 1989; 32(10): 366-9.**

Since its introduction by Barrows in 1964, the simulated patient method has been used in medical education and for research and assessment purposes in medical practice. Inter-reliability in the use of simulated patients for assessment purposes has not yet been satisfactorily investigated. This yields also to what specifically can be concluded from these assessments. The method is however well documented for educational purposes and for registration of actual procedures in medical practice and its use in these situations is justified. The simulated patient method offers the possibility of more closely investigating inter-doctor variation in general practice.

**Key words** Family practice; Simulated patient.

**Correspondence** J.J. Rethans, University of Limburg, P.O. box 616, 6200 MD Maastricht, The Netherlands.

## Chapter 5: Competence and performance: two different concepts in the assessment of quality of medical care

Reprinted from **Family Practice**, 1990, 7, 168-174

# Competence and Performance: Two Different Concepts in the Assessment of Quality of Medical Care

JAN-JOOST RETHANS,* YVONNE VAN LEEUWEN,† RIET DROP,‡ CEES VAN DER VLEUTEN§ AND FERD STURMANS**

Rethans J-J, van Leeuwen Y, Drop R, van der Vleuten C and Sturmans F. Competence and performance: two different concepts in the assessment of quality of medical care. *Family Practice* 1990; **7**: 168–174. In the debate about 'what is a competent general practitioner?' little attention has been paid to the actual practice situation of general practitioners. This paper, based on the 18 most important studies in the literature about medical competence, tries to re-initiate this debate by proposing a clear distinction between 'competence', (what a physician is capable of doing) and 'performance' (what a physician does in his day-to-day practice). With this distinction we looked at whether studies defined both competence and performance, how they dealt with these concepts, what measurement instruments were used and what the conclusions of the studies were. Although it is the common reasoning that competence is a good predictor of performance this concept could not be affirmed. This survey shows that the majority of studies use wrong concepts and come to invalid conclusions. With the empirical distinction between competence and performance however, this paper proposes new directions for the quality assessment of general practitioners.

The British government's white paper on primary health care[1] has again raised the debate about the quality of care delivered by general practitioners. In this report indicators are mentioned in relation to high quality, such as personal availability, preventive activities, attendence at post-graduate courses and certification procedures. Also, a recent series in the *British Medical Journal*, 'What is a good GP?', poses several questions regarding the criteria against which general practitioners should be measured. The discussion is closely related to the issue 'What is a competent doctor?', 'standard measurements of quality' and to assessment procedures in general practice.[2]

What is a competent doctor? Many authors and organizations in areas of medical research and medical education have struggled with this question. With the introduction of standards and logical branching flow charts in medical practice the debate about 'what is competence', 'what is good', and 'what is quality' is also of current interest in this area. The relevant literature includes a large vocabulary, eg: ability, behaviour, performance, clinical judgment, clinical reasoning, habitual performance, problem-solving, clinical competence and the combination of knowledge, skills and attitudes.

The fundamental problem of this lack of clarity about competence is that each author means different things when using the word 'competence'. Other authors deal with the same issue but use different names. The use of different instruments to measure competence, for instance 'paper and pencil tests', 'chart audit', and 'simulated patients', complicates the issue further.

This paper intends to give a clearer picture of 'competence' and specifically proposes clear distinction between it and 'performance'. The reason for doing so is because we think that the debate about the quality of medical care will gain in clarity if this distinction is used. After making the distinction between 'competence' and 'performance' this will be applied to a synthesis of the literature examined from several viewpoints: to see if studies define both competence and performance, how the studies deal with these concepts, what measurement instruments were used and finally the conclusions of the study. Details of the psychometric characteristics of different measurement

*Department of General Practice, †Department of General Practice, ‡Department of Medical Sociology, §Department of Educational Development and Research, and **Department of Epidemiology, University of Limburg, Postbus 616, 6200 MD, Maastricht, The Netherlands. Correspondence to: Dr Rethans.

instruments will not be examined. Finally this paper indicates directions in which assessment of the quality of care delivered by general practitioners might go. This paper is not intended as a quantitative review of the quality of care literature but it takes account of the 18 important studies on this subject (Table 1).

*The distinction between competence and performance*

Quality of care is the crux of every health care system. In trying to guarantee this quality licensure boards have set up examination systems for medical students, who on passing change from a potential to a real doctor. The reasoning underlying these examinations is that passing them predicts the performance of a doctor in the period after the medical school. In other words passing the examinations predicts quality of care.

Senior[3] and Lloyd[4] distinguished between 'competence' and 'performance'; the former meaning 'what a physician is capable of doing' and the latter 'what a physician actually does in his day-to-day practice'. These definitions clearly demarcate the examination: before and during the examination the candidate must behave to the best of his abilities; he deals with the competence setting. After passing the doctor deals with normal day-to-day situations and attempts to cope with them; he deals with the performance setting.

One might argue that 'competence' is made up of several aspects and that it would be more appropriate to speak of 'competences', referring to competences as a series of parameters such as diagnostic and therapeutic skills, interpersonal skills, etc. In this paper however we want to emphasize the construct of competence as a single entity. With the measurement of competence (or performance) both concepts indeed can be operationalized into several aspects.

The distinction between competence and performance raises the question of their relationship. In this there are two possible approaches.

The first one, a more psychometrical approach, considers competence as the limit of one's performance. Or, stated differently, performance is an indicator of competence and *vice versa;* performance can be observed whereas competence can only be inferred. In this concept competence and performance are not two separate concepts but are separate measurement points of the same concept. An example of this reasoning is the current examination system: passing predicts performance.

In the second approach competence and performance are considered as two separate concepts although they can be related in a very delicate way. This approach takes more account of the actual practice experiences of gneral practitioners. When treating patients doctors are influenced by motivational and situational variables. This situation is completely different from their examination situation.

The first reasoning is the one that most educationalists have followed and is the basis of all existing examination systems.

Whether competence is indeed a good predictor of performance has been studied in eight of the eighteen studies mentioned in Table 1 (studies 1–6, 9 and 10); all but one fail to establish this relationship. Only Peterson concludes that years of experience is an indicator of competence (Study 3). With these results at hand researchers conclude that their measurement instruments are not good enough. They conclude that the reason for not establishing a relationship between competence and performance is a psychometric problem.

As in the past 10 years there is only one direction in which this educational research line will go: to construct more and refined measurement instruments which try to assess competence as closely as possible.

The second reasoning, to consider competence and performance as two different concepts, could be more fruitful. If it could be proved that we have to consider competence and performance as different concepts this could give new impulses for research into the daily work of physicians. More research could be focused then to the diagnostic pathways which doctors use during contacts with real patients and which most times do not seem to be very logical or fail to reach ideal standards.[5]

In order to explore this second reasoning the afore mentioned 18 studies are reviewed to look at how authors dealt with competence and performance. We will adopt the definitions of competence and performance by Senior and Lloyd and apply these to the literature.

*Definitions*

Only three of the studies in Table 1, give an explicit definition of either competence or performance or both (12,13 and 16). Fabb states that 'competence is to assess performance' (13), while Morgan sees a competence level as necessary for adequate performance (12). Only Neufeld explicitly separates both concepts in their definitions using the same theoretical concepts as we do in this paper (16). The 15 other studies do not define competence or performance other than in terms of a measurement instrument. Careful examination of these 15 studies, however, shows that most authors implicitly use different concepts of performance and/or competence.

Because of this implicit use of concepts the assessment of these concepts had to be done in two steps. First we looked how the authors used the terms performance/competence and whether they separated both words or used them alternatively. Only one of the 15 studies implies the competence concept (3). Of the 14 other papers eight use the performance concept only (4, 5, 7, 8, 14, 15, 17, 18). Six papers use both concepts synonymously (1, 2, 6, 9–11), (Table 1, subheadings 4–6).

In the second part of this assessment our definitions of performance and competence were applied to all the studies to see whether these were in accordance with

TABLE 1 *The discussed studies*

| | 1. Hubbard JP *et al.* An objective evaluation of clinical competence. *NEJM* 1965; **272:** 1321–28. | 2. Williamson JW. Assessing clinical judgment. *J Med Educ* 1965; **40:** 180–7. | 3. Peterson OL. Attributes of life-long competence in medical practice. *JAMA* 1966; **198:** 765–6. | 4. Gonella JS *et al.* Evaluation of patient care. *JAMA* 1970; **214:** 2040–3. | 5. Barro AR. Survey and evaluation of approaches to physician performance measurement. *J Med Educ* 1973; **48:** 1051–93. | 6. Wingard JR *et al.* Grades as predictors of physician career performance. *J Med Educ* 1973; **48:** 311–22. |
|---|---|---|---|---|---|---|
| 1. | **Purposes of study** To determine the format of the examination of medical candidates | To determine an index of competence | To describe the state of the art | To measure knowledge and to compare that with actual treatment | To investigate the dimensions of performance and how they have been measured | To see if there is a relation between the score of pass/fail students and their career |
| 2. | **Definition of competence** no (see aspects) | no | no | no | no | no |
| 3. | **Definition of performance** no | no | no | no | no (see aspects) | no |
| 4. | **Implicit use of competence** no | no | yes | no | no | no |
| 5. | **Implicit use of performance** no | no | no | yes | yes | no |
| 6. | **Implicit use of comp/perf alternating** yes | yes | no | no | no | yes |
| 7. | **Use of competence according to senior** yes | yes | no | no | no | yes |
| 8. | **Use of performance according to senior** no | no | yes | no | no | yes |
| 9. | **Use of comp/perf alternating according to senior** no | no | no | yes | yes | no |
| 10. | **Indicators of comp/perf** – | – | Years of experience is an indicator of competence | – | – | – |
| 11. | **Discussed aspects of competence** Nine major areas: history, phys.ex/tests & procedures diagnostic acumen, treatment, judgement & skill in implementing care, continuing care, doctor–patient relationship, responsibilities as doctor | Two separate components efficiency, measured from doctors performance; proficiency, measured from the results of this performance in a patient | – | – | – | – |
| 12. | **Measurement instruments** Multiple choice question Written format | Written format (PMP) | – | Chart audit Multiple choice | | – |
| 13. | **Standards discussed** – | External critera by experts (all N–232<60%) | – | Items scored by referee panel. Only 40% of criteria is met | – | – |
| 14. | **Results/conclusions of the study** The 3 measurement instruments are good for assessing competence, but have no predictive validity | Findings illustrate varied and indistinguishable levels of performance quality of docs, irrespective of specialty and certification. There is a negative correl. between performance and years of experience. Diagnostic proficiency decreases; efficiency stays equal | Probability of competence increase with amount training. Most neglected aspect of continuing competence is volume of experience. Measured academic ability has no relation with competence | Performance on MC is not predictive of performance in actual care. It is not a matter of correcting deficiency of knowledge but of translating knowledge into action | At present there exists no system for measuring the overall performance of doctors that has been validated in the sense that those who measure higher have been shown to produce better patient outcome | There is no correlation between grades and career. Actual behaviour is different from ideal behaviour |

TABLE 1 *Continued*

| 7. Fessel JW *et al.* Assessing quality of care from the medical record. *NEJM* 1972; **286:** 134–8. | 8. Lyons TF *et al.* The relationship of physicians' medical recording performance to their medical care performance. *Med Care* 1974; **12:** 463–9. | 9. Mcgaghie WC *et al.* Competency based curriculum development in education. Geneva WHO 1978. | 10. Page GG *et al.* Performance on PMP and performance in practice: are they related? *J Med Educ* 1980; **55:** 529–37. | 11. Norman GR *et al.* A comparison of behaviour on simulated patients and PMP's. *Med Educ* 1981; **15:** 26–32. | 12. Morgan MK *et al.* Evaluation of clinical competence in the health professions. CV Mosby Company, Saint Louis, 1978. |
|---|---|---|---|---|---|
| 1. **Purposes of study** | | | | | |
| To see if the recorded process of medical care and outcome are related | To report correlations between 2 types of quality of care: those that depend upon doctors recording and those that are independent of that | To describe competence in medical education | To test criterion validity of PMP by direct comparison with actual practice | To compare performance of students on PMP and on SP | To improve procedures for evaluating students performances in hospitals laboratory and health care clinics |
| 2. **Definition of competence** | | | | | |
| no | no | no (see aspects) | no | no | The knowledge and skills that are necessary for adequate performance in the profession |
| 3. **Definition of performance** | | | | | |
| no | no | no | no | no | Naturalistic behaviour |
| 4. **Implicit use of competence** | | | | | |
| no | no | no | no | no | yes |
| 5. **Implicit use of performance** | | | | | |
| no | yes | no | no | no | yes |
| 6. **Implicit use of comp/perf alternating** | | | | | |
| no | no | yes | yes | yes | no |
| 7. **Use of competence according to senior** | | | | | |
| no | no | no | yes | yes | yes |
| 8. **Use of performance according to senior** | | | | | |
| yes | yes | no | yes | no | yes |
| 9. **Use of comp/perf according to senior** | | | | | |
| no | no | yes | no | no | no |
| 10. **Indicators of comp/perf** | | | | | |
| – | Good medical care keeping is indicator of good performance | Performance on PMP must be performance in real practice | Naturalistic behaviour versus Performance test | – | – |
| 11. **Aspects of competence** | | | | | |
| – | – | 1. Analysis of docs activities—(self report, observation, test analysis)<br>2. Critical elements of behaviour (critical incident, expert judgment)<br>3. Health care needs (health statistics, medical records, social, economic, pol. realities<br>4. Professional performance model | – | – | – |
| 12. **Measurement instruments** | | | | | |
| Chart audit | Chart audit | See aspects | PMP Simulation in practice | PMP SP in laboratory | All methods |
| 13 **Standards discussed** | | | | | |
| – | Predetermined criteria for optimal care, by experts | Competence is not a matter of comparison; pass/fail. Absolute standards required | Standard is item selected by experts. Only 40 to 50% is being met | Selected options compared with experts | Competence should ideally be objectives. These are specific knowledge and skills the doctor is expected to master |
| 14. **Results/conclusions of the study** | | | | | |
| Neither quantity, nor quality of recorded items is correlated to diagnosis | Correlation between good medical care and recording is +.23 | Definition of competence is bound to local, political, social, economic circumstances, to health needs, resources and structure of health system | The PMP is a good predictor of what will not be done in actual practice | The PMP has a large, significant effect on the number of options selected | There is a difference between a performance-test and habitual behaviour |

| 13. Fabb WE *et al.* The assessment of clinical competence. MTP Press Limited, Lancaster, 1983. | 14. Kemeny ME *et al.* Measuring adequacy of physician performance. *Med Care* 1984; **22:** 620–31. | 15. Sanazaro PJ *et al.* Measuring clinical performance of individual internists in office and hospital practice. *Med Care* 1985; **23:** 1097–114. | 16. Neufeld VR *et al.* Assessing clinical competence. Springer Publishing Company, New York, 1985. | 17. Norman GR *et al.* A comparison of resident performance on real and simulated patients. *J Med Educ* 1982; **57:** 708–15. | 18. Norman GR *et al.* Measuring physicians performances by using simulated patients. *J Med Educ* 1985; **60:** 925–34. |
|---|---|---|---|---|---|
| 1. **Purposes of study**<br>To make an examination handbook which described methods in current use for measuring competence in general practice | To compare the content-validity, scorability, costs and acceptability of 4 methods in measuring adequacy of performance | To test an approach that reliably measures the clinical performance of doctors against national standards | To describe several tests to measure competence | 1. There is no difference in the performance of residents with real or simulated patients in data gathering.<br>2. Residents will be unable to distinguish simulated from real patients | 1. To examine the validity of several methods of criteria setting.<br>2. Difference between criteria and non-criter. docs?<br>3. The reliability of docs perf. scores across patient problems |
| 2. **Definition of competence**<br>To measure competence is to measure performance by assessing behaviour in knowledge, skills, attitudes | no | no | refer to Senior | no | no |
| 3. **Definition of performance**<br>no | no | no | refer to Senior | no | no |
| 4. **Implicit use of competence**<br>– | no | no | – | no | no |
| 5. **Implicit use of performance**<br>– | yes | yes | – | yes | yes |
| 6. **Implicit use of comp/perf alternating**<br>yes | no | no | no | no | no |
| 7. **Use of competence according to senior**<br>no | no | no | yes | yes | no |
| 8. **Use of performance according to senior**<br>no | no | yes | yes | no | yes |
| 9. **Use of comp/perf alternating according to senior**<br>yes | yes | no | no | no | no |
| 10. **Indicators of comp/perf**<br>– | – | For reasonable able docs performance on few cases cannot serve as indicator of performance for others | Knowledge could be predictor of competence as experience. Content specificity is large problem | – | Increasingly apparent that accurate assessment of competence cannot be infered from performance on few problems. Repeated sampling is necessary |
| 11. **Aspects of competence**<br>1. Understanding the individual, family & community<br>2. Analyses and defining health problems<br>3. Managing health problems<br>4. Preventive approach to health care<br>5. Accepting responsibility as a doctor | – | – | 1. Clinical skills<br>2. Knowledge & understanding<br>3. Interpersonal skills<br>4. Problem solving & clinical judgment<br>5. Technical skills | – | – |
| 12. **Measurement instruments**<br>All methods discussed | Physician interview; Patient interview; Videotaped consultation; Chart audit | Chart audit | All methods | (Simulated) patients in laboratory | Simulated patients in real practice |
| 13. **Standards discussed**<br>They desire criterion related standards | Use of criteria maps. Only 50% is met | Peer review with current National standards. 15% of docs score substandard | – | Expert criteria. Only 50% is met | Expert criteria. 65% was met. Only 50–60% by criteria setting doctors |
| 14. **Results/conclusions of the study**<br>Test selection should be Essay 6%; Case 6%; MC 12%; Clinical interpretation 10%; PMP 16%; Diagnostic interview 24%; Physical examination 10% | None of 4 methods is best to measure all aspects of adequacy of doctors' performance. A combination is suggested | Performance is inversely related to years of practice. It would help if could be shown that scores on tests could be related to actual practice. Recertification could raise standards of doctors in real practice | Two decades ago: competence is knowledge. Now competence is multidimensional | The observed behaviour of a doctor with a simulated patient in a laboratory may bear little resemblance to day-to-day activities of doctors | The discrepancy between ideal care and doctors performance probably reflects a lack of realism in setting performance criteria more than it reflects poor performance |

those of the authors. These results are presented in the subheadings 7–9 of Table 1.

It appears that in 10 of the 18 papers the definitions described by the authors are at variance with ours (1–6, 10, 11, 14, 17). With this result at hand we conclude therefore that four studies describe only competence (1, 2, 11, 17). Five studies deal with performance (3, 7, 8, 15, 18). Another four papers deal with both competence and performance but separate these concepts in their study (6, 10, 12, 16). The remaining five studies use both concepts alternatively and without distinction (4, 5, 9, 13, 14).

*Use of measurement instruments*

Inherently related to the use of competence/performance concept is the use of measurement instruments. It is argued here that the distinction between competence and performance has its implications for the use of measurement instruments. If the object of study (competence, performance) and used measurement instrument do not match, then research conclusions are invalidated.

Measurement instruments (or methods) can be divided into direct and indirect methods. With direct methods the research workers see or hear a physician dealing with patients or with examiners. Such methods include observation of doctors (both video and audio), use of standardized patients or oral examinations. With indirect methods direct observation of doctor–patient contacts is not possible. They depend upon written simulation papers, chart audits, interviews and multiple choice questions. Using the concepts of competence and performance it is clear that some instruments may be used in both concepts, while others may not. Written simulation papers, laboratory observation, oral examination, multiple choice questions can only be used in the competence-situation. Chart audit can be used for performance measurement. One of the best methods for performance measurement (high reliability and validity) is some kind of unobtrusive measurement such as the use of standardized patients in a blinded way, so that doctors do not know when they are dealing with such a patient. We analysed 16 of the 18 studies (3 and 6 are review papers) to see whether the appropriate measurement instrument had been used. For example if an author used a written simulation method and yet labelled the study as 'measuring performance' this is considered incorrect. Five (31%) used incorrect instruments (2, 5, 9, 14, 17) or labelled their instruments incorrectly. In these five studies the authors' purposes were to describe performance, but instead they measured competence.

## DISCUSSION

This paper shows that research workers do not specify very carefully which concept of competence/performance they mean. One might argue that authors do not give definitions because it is neither important nor necessary. However because every author use different concepts implicitly, a definition is imperative and enables the reader to interpret the situation he is dealing with.

The comparison between the authors' implicit definitions of performance/competence and the way we classified these according to our concepts revealed a disturbing picture. It is disappointing that so many papers fail to delineate what they are dealing with and it is even more disturbing that most authors deal with concepts which they label incorrectly. The consequence of this misclassification is also shown in the use of measurement instruments: Even among the 11 studies in which the right instruments were used four authors interpreted their own concept falsely (1, 4, 10, 11). This means that in only seven (44%) studies (7, 8, 12, 13, 15, 16, 18) the right instrument matched the right concept!

We are inclined to consider only the conclusions of these seven studies as valid. Fessel (7) and Lyons (8) conclude that they did not find a significant relation between chart audit and good medical care. Fabb (13) and Neufeld (16) give a frame work for a test to measure competence. The other studies, Morgan (13), and Sanazaro (15) and Norman (18), refer to a possible distinction between competence and performance. This is in line with our reasoning. It is clear that the majority of studies adhere to the reasoning that competence is an indicator of performance. But, as we stated earlier, all but one failed to establish this relationship.

To develop a better empirical basis for the debate about the quality of medical care future studies should concentrate on the difference or relation between competence and performance. The implication of the results of these future studies might have great impact both on medical schools and medical practice. Will it still be justified to teach medical candidates 'competence', or should medical schools concentrate more on 'performance'? Sibley, in a study to determine the effects of continuing medical education, concluded that although the participants showed significant gains in their knowledge, they did not change in their performance.[6] Norman showed that doctors in actual practice performed considerably below the criteria developed by themselves.[7] Should standards of quality of medical care, delivered by doctors in their daily practice, continually be used based on 'competence' standards?

To find an answer to these questions we think the recommendations are essential. Firstly, researchers should make a clear distinction between 'competence' and 'performance'. We think this could be done according to our definitions. Secondly when using measurement instruments research workers should choose their instrument in correspondence with the object of study. In general it can be stated that different concepts require different measurement instruments. For those instruments which seem appropriate for both competence and performance settings researchers

should carefully describe in which setting they are used during the study. At present we are running a study in which we try to find more empirical evidence for the distinction between competence and performance. We compare general practitioners in three settings. In the first setting doctors will be visited by standardized patient, indistinguishable from real patients. In the second setting the same doctors will deal with the same standardized patients in a laboratory situation (video controlled). Finally the doctors will handle a written simulation about the same patients as they met before.

Other research questions in this area are whether a relation between performance and output of medical care exists. What factors influence the performance of doctors and how may they be influenced to raise the standards of performance?

It is our conviction that the debate about the quality of medical care can take place at a higher level if these fundamental questions are answered clearly.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Robin Hull (Birmingham), Pie Hobus and Wim van Zutphen for their comments on earlier drafts of this paper.

REFERENCES

[1] Secretaries of State for Social Services, Wales, Northern Ireland and Scotland. Promoting better Health. London: HMSO, 1987. (Cmnd 249).

[2] Gray DP. Assessment in general practice. *J R Coll Gen Pract* 1988;**38**:344–345.

[3] Senior JR. Towards the measurement of competence of medicine. Philadelphia: National Board of Medical Examiners 1976.

[4] Lloyd JS. Definitions of competence in specialties of medicine. Chicago: American Board of Medical Specialties 1979.

[5] Rethans JJE, Van Boven CPA. Simulated patients in general practice: a different look at the consultation. *Br Med J* 1987;**294**:809–812.

[6] Sibley JC, Sackett DL, Neufeld V, Gerrard B, Rudnick KV, Fraser W. A randomized trial of continuing medical education. *New Engl J Med* 1982;**306**:511–515.

[7] Norman GR, Neufeld VR, Walsh A, Woodward CA, McConvey G. Measuring physicians' performances by using simulated patients. *J Med Educ* 1985;**60**: 925–934.

## Chapter 6: A method for introducing standardized (simulated) patients into general practice consultations

Reprinted from the **British Journal of General Practice**, 1991, 41, 94-96

# A method for introducing standardized (simulated) patients into general practice consultations

JAN-JOOST RETHANS

RIET DROP

FERD STURMANS

CEES VAN DER VLEUTEN

*SUMMARY. A study has been undertaken to determine whether it is possible for a set of standardized (simulated) patients to visit general practitioners, without being detected, in a health care system where doctors have fixed patient lists. Since sending standardized patients into doctors' offices is a new way to assess the performance of general practitioners; this paper describes in detail the methodology that has been used for visits.*

*The paper looks first at the general preparation for visits and secondly at the specific preparation concerning the fine detail of the individual visit. The method was tested in 156 consultations with 39 general practitioners and in no cases were the standardized patients detected. None of the doctors visited felt offended and all were prepared to cooperate in future studies with standardized patients. It is concluded that the standardized patient method, following the step-by-step procedure described, is feasible in actual practice.*

## Introduction

RECENTLY the need has been stressed for new methods of assessing the actual performance of general practitioners rather than assessing what they are capable of doing: their competence.[1,2]

Doctors' behaviour may be measured directly or indirectly. With direct methods the research worker observes the physician dealing with patients by means of video- or audio-tapes, or the use of standardized patients. Indirect methods consist of chart audits or written or oral examination. The choice of a particular method depends on whether one is interested in actual practice of doctors or in test (competence) situations. For both situations high validity and high reliability of a method are essential.

In the last two decades the standardized (or simulated) patient has proved a powerful high quality instrument for assessing the competence of medical students and doctors.[3] This method has been extended by introducing standardized patients, indistinguishable in almost every case from real patients, into general practice. Since the method is a direct one and both reliability and validity are high,[4] it may be the best method of assessing clinical care. Experience of this method in actual practice is limited; only four studies making valid, reliable use of standardized patients in practice have been reported.[4-7] Woodward described the use of standardized patients in the North American situation.[8] In view of the need for more methods for assessing performance in actual practice, and in view of the lack of experience with the use of standardized patients in actual practice, the department of general practice at the University of Limburg has undertaken a study to determine whether it is possible for a set of standardized patients to visit general practitioners without being detected, in a health care system where general practitioners have fixed patient lists. This paper reports in detail on the method that has been used.

The preparation for visits consisted of: first, general preparation in which the whole operation was organized regardless of individual practices or doctors; and secondly, specific preparation on fine detail of the individual visit.

## General preparation for the visits

### *Preliminary mailing of information to doctors*

It was important to explain this research as a means of studying what happens in the consultation so that doctors did not perceive standardized patients to be offensive or intrusive. The purpose was not to look for 'rotten apples' but to show that knowledge of what really happens in the consultation improves our understanding of why doctors behave as they do.[9] Doctors were paid for their care of the standardized patients and all data was treated in strict confidence and analysed blind. Feedback was provided to individual doctors and consisted of an item list, scored by the patients, about the performance of the individual doctor, compared with his or her peers. This procedure ensured that none of the participating doctors experienced ethical problems with this type of research in general practice. Doctors did not know when or how often they would be visited by standardized patients. Information was sent to doctors at the start of preparation thus ensuring a long interval between consent and the visit.

Doctors were also sent 'detection forms' which they were asked to use to report suspected standardized patients, giving the patient's name, the date of the visit and degree of the doctor's certainty of having identified a standardized patient.

### *Selection of simulated medical problem*

Any medical problem without physical signs, and many conditions with physical signs, can be simulated effectively. For example, jaundice can be simulated with make-up. Standardized patients can also be recruited from real patients with stable physical pathology, for example heart murmurs.[3] It was our experience that a vivid imagination among researchers was of great value in designing different types of simulations.

There are four possible types of consultations made by standardized patients: (1) a patient consulting while on holiday; (2) a consultation outside practice hours while the doctor is deputizing for others; (3) a consultation during normal working hours by a patient working in the neighbourhood with an acute problem (for example chest pain or foreign body in the eye); (4) a first consultation by a patient who has been newly enlisted in the practice.

An additional variant is the home visit but this has not been used here.

### *Setting up background data*

Many general practitioners will ask the name of a patient's previous doctor. Therefore standardized patients were provided with names of doctors who had been briefed with written descriptions of the standardized patient's role so that they could respond to a request from the other doctor.

Jan-Joost Rethans, MD, assistant professor, Centre for Quality Assurance in Research in General Practice, Universities of Nijmegen and Maastricht, The Netherlands; Riet Drop, professor, Department of Medical Sociology; Ferd Sturmans, professor, Department of Epidemiology; Cees van der Vleuten, PhD, associate professor, Department of Educational Development and Research, University of Limburg, Maastricht, The Netherlands.

Submitted: 16 May 1990; accepted: 4 September 1990.

*British Journal of General Practice*, 1991, **41**, 94-96.

Seventy per cent of patients in the Netherlands are insured for medical care through health insurance schemes or *ziekenfonds* and must show their insurance cards to their doctor at each visit. These insurance cards show the private health insurance number, name, address, date of birth and name of the patient's general practitioner, pharmacist and dentist. To keep detection of the patients as low as possible the *ziekenfonds* were asked to cooperate by supplying the standardized patients with real insurance cards (new patient names) and to enlist these patients on the lists of the doctors: this also allowed doctors to be paid for their consultations. For the first three types of consultation this insurance card procedure was not really essential, but it was necessary in order to get a patient newly enlisted in the practice. During the project the patients remained on the list of their own real general practitioner in case of genuine illness.

### *Selection of standardized patients*

Any motivated person may become a standardized patient and many medical schools have pools of standardized patients available for educational purposes.[3] The standardized patients used in our studies were selected from a pool of 100 in the skills laboratory of the medical school at Maastricht. They were paid for their participation. The criteria used in selection were that standardized patients matched the roles assigned to them (a 40-year-old woman cannot be simulated by an 18-year-old man); were able to memorize 20–40 items of information; were able to cope with stressful events; ideally possessed a driving licence and were independent of public transport (it is unconvincing to be very late for an appointment when the patient supposedly lives close to the practice).

Standardized patients were required to make a written undertaking to keep all information about doctors strictly confidential.

Though some standardized patients were able to visit up to 20 doctors in a given role most managed 13 and it was found very useful to have some reserves in case of real illness.

### *Role and reliability training*

The standardized patients were trained to play their role as a patient and to report reliable and valid facts about the consultation. This training took place at the medical school and lasted about 14 hours. During these sessions the patients repeatedly played their role in contact with other doctors. With the use of videotapes of these sessions the patients were supplied with feedback about how they performed. To check the reliability among the standardized patients three videotaped test consultations were evaluated. In these tests the patients played their role in a contact with a staff member of a university department of general practice. The patients completed a checklist of items, based on a consensus set of standards of care (see the accompanying paper),[7] immediately after such a consultation; the same consultation was also scored by three independent doctors. For each consultation the scores obtained by the three independent doctors were considered to be the gold standard for that particular consultation. Subsequently, the individual scores of the standardized patients were compared with this gold standard to assess their reliability. To assess the consistency of individual patients each patient was retested after about six weeks, after the visits to the doctors by scoring the same consultation recorded on videotape. These procedures have been carried out successfully before.[5,6] The reliability and consistency agreement scores all ranged from 0.9 to 1.0 (kappa 0.8–1.0).

## Specific preparation for a visit

### *Selection of doctors to be visited*

In some cases economic considerations influenced the selection of a doctor from the list of those willing to participate. It was important not to select doctors who could detect the standardized patient, therefore only one doctor per practice was selected. Caution was exercised in selecting doctors from small communities where everybody knows everybody or colleagues of the researchers with possible access to information about the standardized patient's medical problem or date of visit. Standardized patients were not asked to make these assessment visits to their own real general practitioner. They were sent to doctors and practices that they had never been to before, often travelling to other cities. This minimized recognition of the standardized patients by practice staff or other patients.

### *Reconnoitre of practice and district*

All selected practices were visited by the first author to gain knowledge that would make the stories of the standardized patients credible. This enabled information to be collected about possible addresses for the patients, about places where the patients could be employed and about some local interesting buildings or events.

### *Selection of home addresses for standardized patients*

The address used by a standardized patient needed to be credible to the doctor being consulted yet not so familiar that he was capable of recognizing its falsity. Addresses were selected in three ways. It was sometimes possible to access the computer of another doctor with a practice in the same city where the address of a real patient could be 'borrowed'. Real addresses of friends and relatives of the researchers were also used. If these did not succeed a risk was taken in selecting a non-existing house number in a known street. Once names and addresses were selected the doctors were notified by a routine mailing of the *ziekenfonds* of the inclusion of the patient on their list. This notification was made between two and five months before the standardized patient visited the doctor.

### *Information given to standardized patients*

Standardized patients were given detailed information about the practice to be visited, such as telephone number, appointment system, address details of the doctors and other staff, the number of medical secretaries and the other health personnel (physiotherapists, practice nurses) present. Some practices received students or trainees from the research workers' medical school. Where possible standardized patients were told of the experiences of other standardized patients who had made previous visits. It might be argued that most real patients are not usually so well informed about a practice, but it should be borne in mind that standardized patients had to concentrate on their role. Every item of information about practices helped them to feel 'at home' in the practices and enabled them to concentrate better on essential items of their role. Around six hours of additional training was needed to give this information to patients about the practice they were going to visit and for return meetings around the time of and after the actual visits to the general practitioners.

### *Selection of dates for visits*

After studying details about the practices the standardized patients were asked to visit the practices as soon as possible, choosing a date which would ensure that they saw their target doctor rather than a deputy.

### *Pilot visits*

After the training programme standardized patients made a single pilot visit — that is, their first real visit in actual practice as a standardized patient — and this was evaluated at the medical school. These pilot visits were always successful and increased the confidence of the standardized patients.

## Feasibility study

In January 1988 all 442 general practitioners working in the province of our university received information about this study and were asked to give their written permission that they would accept standardized patients into their practices until January 1991. The doctors were not told how many times and when they would be visited, nor were they informed about the content of the medical complaints. The doctors were told that at the end of the project they would receive information about which standardized patients had visited them. Of the 442 doctors, 137 agreed to participate and were sent a 'detection form', which had to be returned immediately after they thought they had detected a standardized patient.

Of the 137 doctors who agreed to participate, 39 were selected. The main criterion for including a general practitioner was that the distance between his or her practice and the university was less than 30 kilometres. This was done for financial reasons. Each selected general practitioner was visited by four different standardized patients presenting four different medical complaints during a four month period starting at least 12 months after they agreed to participate.

None of the standardized patients was detected. Two doctors returned a completed detection form but both forms reported real patients.

## Discussion

The ethical problems of sending standardized patients into doctors' offices were discussed at the start of this project with representatives of the Dutch college of general practitioners (Nederlands Huisartsen Genootschap). It was concluded that there were no ethical problems if doctors gave a written consent to be visited and if they were informed at the end of the project when and by which standardized patients they had been visited.

It was encouraging that in none of the 156 consultations was a patient detected, even though each doctor was visited by four patients. There are probably several reasons for this success. First, there was a lengthy period between the doctors' consent to participate and the actual visits. Secondly, the original health insurance cards certainly helped the patients to validate their position. In at least two visits the patients thought that the doctors were suspicious at the start of the consultation. In both cases the original insurance papers convinced the doctors that nothing was wrong. Thirdly, the standardized patients were supplied with a great deal of inside information about the practice they were going to visit. Thus, before a particular visit, the only thing the standardized patient did not know was what the doctor looked like. During the project the standardized patients also learned to adopt strategies for dealing with the doctors' secretaries. Although we had thought that it would be difficult for standardized patients to visit doctors in small villages without detection, all eight such consultations succeeded. The selection of the practices may perhaps have been too strict. It can be concluded that for assessment purposes the standardized patient method is a feasible and adequate method, even in a health care system where doctors see only patients who are registered in their practice.

There are some additional aspects of standardized patient studies which are worth mentioning. The method was not expensive: the budget of this project (lasting two years) was calculated to be about £32 300. This sum includes the payment and training of the patients involved (£1100) and the fees for the participating doctors (£320). The rest of the budget consisted of the salary for the researcher (one doctor working four days a week).

Many non-medical issues were reported spontaneously by the standardized patients. For example, at the beginning of their visits most standardized patients felt embarrassed to find that the doctors were really interested in them. Some doctors used to explain where the patients could find the closest pharmacy or started small talk with the patients, after the complaint had been treated. Other doctors actively helped standardized patients to dress or undress. Some doctors sent written patient information about complaints to the addresses of the patients. Some doctors, after receiving the routine mailing list of new patients, sent an invitation to the patient to come to the practice to get acquainted. Some peculiar details were reported: one doctor started to sing for his patient and another doctor smoked during the consultation. Furthermore, each patient saw many waiting rooms and different styles of interior. The waiting times differed considerably. In some cases standardized patients felt satisfied with the doctor, but not with the practice nurses. Some patients thanked us after the study for enabling them to get this experience and spoke of 'those nice, kind and dear doctors'.

At the end of the studies we informed the participating general practitioners about the methods used. None of the doctors felt offended and all were prepared to cooperate in future studies with standardized patients. All participating general practitioners expressed their belief in research by means of standardized patients in real practice: they considered the data to be valid and very useful for feedback purposes.

We conclude that the use of standardized patients is a method which has proved feasible in actual practice. The limitations of the method are in the simulation of particular medical problems, as we pointed out earlier. Roles in which patients would be at risk of undergoing invasive investigations could also be a problem. In general practice, however, this rarely occurs and if patients are referred to hospitals to undergo such investigations, the results can be simulated too. Until now an important aspect of this type of study was to test whether patients would succeed in entering the practices. This is not a problem and more attention may be paid to more important aspects such as the content of consultations and personal working styles of doctors. We believe that the standardized patient method has enormous potential for research and audit because it gives more insight into what goes on in the consulting room of doctors. One example would be a study in which doctors were asked immediately after a visit by a standardized patient why they acted as they did. From this we could learn more about why doctors act as they do and evaluate how they provide their care.

## References


1. Maguire P. Assessing clinical competence. *Br Med J* 1989; **298:** 4-5.
2. Newble DI. Assessment of clinical competence: state of the art. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP (eds). *Teaching and assessing clinical competence.* Groningen: Boekwerk Publications, 1990.
3. Barrows HS. *Simulated (standardized) patients and other human simulations.* Chapel Hill, NC: Health Sciences Consortium, 1987.
4. Norman GR, Neufeld VR, Walsh A, *et al.* Measuring physicians' performances by using simulated patients. *J Med Educ* 1985; **60:** 925-934.
5. McClure CL, Gall EP, Meredith KE, *et al.* Assessing clinical judgment with standardized patients. *J Fam Pract* 1985; **20:** 457-464.
6. Rethans JJE, Van Boven CPA. Simulated patients in general practice: a different look at the consultation. *Br Med J* 1987; **294:** 809-812.
7. Rethans J-J, Sturmans F, Drop R, van der Vleuten C. Assessment of the performance of general practitioners by the use of standardized (simulated) patients. *Br J Gen Pract* 1991; **41:** 97-99.
8. Woodward CA, McConvey GA, Neufeld V, *et al.* Measurement of physician performance by standardized patients. *Med Care* 1985; **23:** 1019-1027.
9. Berwick DM. Continuous improvement as an ideal in health care. *N Engl J Med* 1989; **320:** 53-56.



**Acknowledgements**

We thank the General Practitioners Writers Association and in particular Robin Hull for their help in 'polishing' this paper.



**Address for correspondence**

Dr J-J Rethans, Department of General Practice, University of Limburg, Postbus 616, 6200 MD Maastricht, The Netherlands.

## Chapter 7: Assessment of the performance of general practitioners by the use of standardized (simulated) patients

Reprinted from the **British Journal of General Practice**, 1991, 41, 97-99

# Assessment of the performance of general practitioners by the use of standardized (simulated) patients

JAN-JOOST RETHANS

FERD STURMANS

RIET DROP

CEES VAN DER VLEUTEN

*SUMMARY. A study was undertaken whereby a set of standardized (simulated) patients visited general practitioners without being detected, in a health care system where doctors had fixed patient lists. Thirty nine general practitioners were each visited during normal surgery hours by four standardized patients who were designed to be indistinguisable from real patients. The objective of the study was to see whether the actual performance of general practitioners, as assessed by standardized patients, met predetermined consensus standards of care for actual practice. The patients presented standardized accounts of headache, diarrhoea, shoulder pain and diabetes. The mean group scores of the doctors on the predefined standards of care for the different complaints ranged from 33 to 68%. The results show that standardized patients may be the method of choice in the assessment of the quality of actual care of doctors. It is hypothesized that the substandard scores of the doctors do not reflect inadequate competence, but are a result of the difference between competence and performance.*

## Introduction

IN the current debate about the quality of the performance of general practitioners, problems arise in defining the methods for assessing quality of practice.[1] Good methods are those which possess high validity and high reliability, but in real practice the feasibility of a method is also an important aspect. Traditional assessment methods have relied on written tests and clinical examinations but doubts have been cast on their validity and reliability.[2] Audit of medical records has been criticized for its low reliability in assessing several aspects of a consultation.[3,4] Audio- and video-taping of consultations are methods which have both high validity and reliability. A disadvantage of these methods is that the researcher cannot control which patients enter the surgery room, making it difficult to compare performance between doctors. This is not a problem with the standardized (or simulated) patient method and therefore this method has been described as the best one for assessing the management of patients by doctors.[4] With the help of standardized patients it has recently been shown that doctors in real practice performed more actions categorized as essential for good quality care than they said they would in an open ended questionnaire.[5] Experience with the use of standardized patients who report data which can be considered reliable and valid is, however, highly limited and requires further testing.[4-6]

Problems arise not only in defining the methods of assessing the performance quality of doctors, but even more in defining the level of quality of care, in other words, deciding what is 'competence', what is 'good' and what is 'bad'. One of the most common ways to define a level of quality for performance is to set standards for actual health care. In most instances a group of experts in general practice determines a set of standards for a particular medical problem. This procedure, however, carries the risk of setting 'armchair standards', or standards that have no basis in actual practice.[7] It has, for example, been shown that a group of doctors who were first asked to assess a practical standard for a particular medical complaint, performed at only 56% of their own standard during actual consultations.[4] The department of general practice at the University of Limburg has undertaken a study to determine whether the actual performance of general practitioners, as assessed by standardized patients, meets predetermined standards for actual practice.

## Method

From among 24 nationally accepted sets of standards of care in general practice in the Netherlands, eight medical problems were identified which were all common in general practice, presented a diagnostic challenge and could be presented by a standardized patient.[8] The 24 sets of standards had been determined by a consensus procedure with several stages. During the stages, experts in general practice as well as general practitioners 'in the field' commented on the sets of standards and tested them in practice. It was stressed that the standards should be practicable for actual practice and that they should not reflect an 'academic view' of real practice. The standards were divided into three categories: essential actions (considered to be necessary for good quality care), intermediate actions (not essential for good care but not harmful either) and superfluous actions. The eight problems which were selected for the study were translated into standardized roles for standardized patients to use. A panel of three general practitioners independently ranked these roles with respect to face validity. The four cases which ranked highest and on which there was agreement among the panel were finally chosen for this study. Figure 1 shows the most important features of the four cases. The facts which the doctor collected about history, the results of physical and laboratory examination, instructions given to the patient, treatment, and follow up were to be reported by the standardized patients and scored according to the sets of standards. The number of items on each standard ranged from 26 to 36. The standards used and their preparation process have been published in detail before.[8] Figure 2 represents an example of one of the standards used.

Twelve standardized patients (six women and six men) were selected. Three standardized patients were allocated to each case study; these were of the same sex and approximately the same

Jan-Joost Rethans, MD, assistant professor, Centre for Quality Assurance in Research in General Practice, Universities of Nijmegen and Maastricht, The Netherlands; Ferd Sturmans, professor, Department of Epidemiology; Riet Drop, professor, Department of Medical Sociology; Cees van der Vleuten, PhD, associate professor, Department of Educational Development and Research, University of Limburg, Maastricht, The Netherlands.

Submitted: 16 May 1990; accepted: 4 September 1990.

| | |
|---|---|
| Case 1. | A 30-year-old man visits his general practitioner with a headache. He is to say that he is an assistant accountant and has had the headache since he started his new job, three months ago, at an accounts office. The pressure to succeed at his job is quite considerable. In addition he has had to move away from his girlfriend, who now lives 200 kilometres away and whom he can meet only at the weekends. |
| Case 2. | A 26-year-old woman presents to her general practitioner with acute diarrhoea of three days onset and asks for something to stop her complaint. If the doctor asks questions about her job situation, she is to say that she works at a butcher's shop. |
| Case 3. | A 40-year-old woman presents to her general practitioner with pain in one shoulder, which she has had for four days. She is newly divorced and has a daughter aged 15 years. Should the doctor ask, she is to answer that the pain started after she had painted several doors in her new apartment. |
| Case 4. | A divorced man, 59 years old, with a diabetes type 2, visits his general practitioner for a new (repeat) prescription for his oral antidiabetic medication. |

**Figure 1.** *Features of the four standardized medical complaints.*

*History*

1. Essential. Ascertain presence of diarrhoea.
2. Essential. Ascertain presence of nausea, vomiting.
3. Essential. Ascertain presence of abdominal pain.
4. Essential. Ascertain course and duration of the complaints.
5. Essential. Ask for possible causes such as contact with illness, travel, special foods etc.
6. Essential. Ask whether any other complaints are present.
7. Intermediate. Check medication being taken.
8. Intermediate. Check eating habits history.
9. Intermediate. Check for fever.
10. Essential. Ascertain stool consistency and frequency.
11. Essential. Ascertain nature of pain.
12. Essential. Ascertain location of pain, shifts in location.
13. Superfluous. Ask for additional information.

*Physical examination*

14. Essential. Examine abdomen: inspection, percussion, auscultation, palpation.
15. Superfluous. Carry out further examinations.

*Laboratory tests*

16. Essential. Take stool culture.
17. Superfluous. Carry out other laboratory tests.

*Guidance and advice*

18. Essential. Present the diagnosis.
19. Essential. Discuss the prognosis.
20. Essential. Give advice about diet.
21. Superfluous. Give other guidance and advice.

*Medication*

22. Intermediate. Prescribe antiemetics.
23. Intermediate. Prescribe absorbents.
24. Superfluous. Prescribe other medication.

*Return visit*

25. Essential. Explain that patient should return in two or three days if complaints persist.
26. Superfluous. Request return visit.

**Figure 2.** *The standard of care for the standardized patient presenting with diarrhoea; the actions taken by the general practitioner are divided into essential actions, intermediate actions and superfluous actions.*

age. The training of standardized patients for playing their role and for reporting reliable and valid facts about the doctor's performance has been described in another paper.[9] Before the actual visits all standardized patients signed a written consent to keep all medical and personal information about the general practitioners in the project strictly for research purposes. The doctors selected for the study were visited by the standardized patients during a four month period starting at least 12 months after they agreed to participate.

## Results

Of the 442 doctors asked to participate 137 (31%) agreed to be visited. Of these 137 doctors, 39 were selected. The criteria for selection have been outlined in a previous paper.[9] The personal and practice characteristics did not differ from national characteristics.

For the subcategories 'essential', 'intermediate' and 'superfluous' actions, Table 1 shows the mean number of actions which were actually performed by the general practitioners for each of the four medical complaints. To some extent scores can be standardized as the percentage of the potential total maximum score in each category of action for the standard. Since the number of superfluous actions can be potentially infinite no percentages were calculated for this category. For each of the complaints, there was a clear distinction in the adherence to the standards between the essential and intermediate actions. We have also analysed which elements of the standards (for example, history or physical examination) were adhered to and which were omitted, but no consistent pattern was found for the four complaints. From Table 1 it can be seen that a much smaller proportion of the maximum scores were obtained by the doctors for the diabetic case than for the other cases.

## Discussion

Since only 31% of doctors approached agreed to participate, the results might reflect performance of the more competent doctors in the province of the medical school. Although the personal and practice characteristics of the participating doctors did not deviate from national characteristics, we cannot exclude the possibility that the performance of non-participating doctors would be different, possibly lower, than the results in this study.

The results showed that in actual practice doctors met only between 33–68% of established national consensus standards for essential actions. Because these standards have been developed for practical use, one would expect that doctors in actual practice would meet 90–100% of the standards. Since 10 visits could be considered the minimum number to be representative of the individual doctor's usual level of performance, the results in this study do not allow us to look at individual doctors,[10] but are only applicable to the participating doctors as a group. The fact that no consistent pattern was found regarding which elements of the standards were adhered to and which were omitted, might be the result of content specificity and of the fact that only four cases were used. An essential question in the current process of standard setting and especially with regard to the substandard scores of the participating doctors is whether this result reflects inadequate performance. In this whole project 137 doctors agreed to be visited over a period of three years. The fact that these doctors were not afraid of being audited with the very direct method of standardized patients, shows that they had faith in their own methods of dealing with patients. The conclusion from this study — that physicians in practice perform at a level considerably below the standards set by their peers — is not new.[4,11] However, this has never before been demonstrated in actual practice with a method as direct as the

Table 1. Number of actions scored (mean, range and interquartile range) for the four medical problems as reported by standardized patients for consultations with 39 general practitioners.

| | Number of actions scored | | | |
|---|---|---|---|---|
| | Mean ($n = 39$) | Range | Inter-quartile range (quart 1 to quart 3) | Mean no. of actions scored as a percentage of maximum no. from standard |
| *Headache case* | | | | |
| Essential actions (maximum 13) | 8.9 | 5–12 | 8–10 | *68* |
| Intermediate actions (maximum 12) | 3.8 | 0–8 | 3–5 | *32* |
| Superfluous actions | 2.6 | 0–7 | 1–4 | |
| *Diarrhoea case* | | | | |
| Essential actions (maximum 15) | 8.5 | 5–14 | 7–10 | *57* |
| Intermediate actions (maximum 5) | 2.4 | 1–4 | 1–3 | *48* |
| Superfluous actions | 1.8 | 0–8 | 0–2 | |
| *Shoulder pain case* | | | | |
| Essential actions (maximum 19) | 12 | 4–16 | 10–14 | *63* |
| Intermediate actions (maximum 6) | 1.7 | 0–3 | 1–2 | *28* |
| Superfluous actions | 1.8 | 0–6 | 1–3 | |
| *Diabetic case* | | | | |
| Essential actions (maximum 21) | 6.9 | 2–14 | 5–8 | *33* |
| Intermediate actions (maximum 6) | 0.2 | 0–2 | 0–0 | *3* |
| Superfluous actions | 2.0 | 0–7 | 1–3 | |

$n$ = number of consultations.

use of standardized patients with four visits per doctor. As already mentioned in the introduction, whether standards are external (as in this study) or internal does not seem to affect the results obtained: doctors still performed below the standards set.

This observation of substandard performance of doctors has in the past led to actions by various professional organizations. These actions were primarily directed towards mandatory attendance at continuing medical education. The underlying assumption of these directives is that poor performance is a reflection of inadequate knowledge and/or skills, which would be remediable by additional instructions. Studies about the effects of postgraduate education on the behaviour of doctors in actual practice often produce conflicting evidence.[12] It has, for instance, been shown that additional postgraduate education does not seem to change the practice behaviour of doctors.[13] Recently it has also been shown in a study of postgraduate teaching of fundoscopy to general practitioners that there was no measureable learning effect, although the doctors were very enthusiastic.[14]

An alternative explanation to that of inadequate competence is that physicians, when not under supervision, do not perform at the level they are capable of.[15] The willingness of all participating doctors to allow their actual behaviour to be scrutinized, however, suggests that the results of this study do not reflect inadequate performance *per se*. The subjective opinions of the standardized patients after their visits support this view. At the first visits in actual practice some patients were almost embarrassed because they felt that the doctors were really interested in them and did their best to help them. The results of this study stress the need for more studies which investigate the relationship between actual behaviour and maximum competence of practising doctors.

Another explanation for the substandard scores of the doctors could be that the standards involved, notwithstanding the consensus procedure used, do not adequately reflect the longitudinal relationship between patients and their general practitioners. It is known that in actual practice doctors show efficient performance and do only what is necessary at that particular moment.[6]

The large difference between the mean score for the chronic diabetic case and those for the other cases could be because doctors do not ask all their questions at the first consultation, but spread their questions over several visits. This hypothesis can be tested by visiting doctors several times with the same standardized patient. Our department of general practice is currently involved in such a study. The results of the present study stress the need to establish standards which take account of longitudinal relationships in general practice.

The finding that doctors perform below predetermined standards does not prove that doctors are incompetent; it should at least be tested against the hypothesis that standards for actual care are still not realistic.

It can be concluded that the standardized patient method is feasible and is the most direct method for assessing the performance of practising physicians. The data gathered with this method about actual practice can give a new stimulus to the debate about what constitutes 'good' and 'bad' in general practice.

## References


1. Mok J. Assessment in general practice *J R Coll Gen Pract* 1988; **38:** 344-345.
2. Maguire P. Assessing clinical competence. *Br Med J* 1989; **298:** 4-5.
3. Fessel JW, van Brunt EE. Assessing quality of care from the medical record. *N Engl J Med* 1972; **286:** 134-138.
4. Norman GR, Neufeld VR, Walsh A, *et al.* Measuring physicians' performances by using simulated patients. *J Med Educ* 1985; **60:** 925-34.
5. McClure CL, Gall EP, Meredith KE, *et al.* Assessing clinical judgement with standardized patients. *J Fam Pract* 1985; **20:** 457-464.
6. Rethans JJE, van Boven CPA. Simulated patients in general practice: a different look at the consultation. *Br Med J* 1987; **294:** 809-812.
7. Neufeld VR, Norman GR. *Assessing clinical competence.* New York: Springer, 1985.
8. Grol R, Mesker P, Schellevis F (eds). *Peer review in general practice.* Nijmegen: University Department of General Practice, 1988: 87-89; 97-98; 109-108; 117-119.
9. Rethans J-J, Drop R, Sturmans F, van der Vleuten C. A method for introducing standardized (simulated) patients into general practice consultations. *Br J Gen Pract* 1991; **41:** 94-96.
10. van der Vleuten CPM. Assessment of clinical skills with standardized patients: state of the art. In: *Naar een rationeel systeem voor toetsing van studieprestaties in probleemgestuurd onderwijs (thesis).* Maastricht University, 1989: 93-140.
11. Gonella JS, Goran MJ. Evaluation of patient care. *JAMA* 1970; **214:** 2040-2043.
12. Haynes R, Davis D, McKibbon A, Tugwell P. A critical appraisal of the efficacy of continuing medical education. *JAMA* 1984; **251:** 61-64.
13. Sibley JC, Sackett DL, Neufeld V, *et al.* A randomized trial of continuing medical education. *N Engl J Med* 1982; **306:** 511-515.
14. Reeders K, Hiemstra RJ, Bender W. Teaching fundoscopy to GPs in an OSCE setting: failures and successes. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP (eds). *Teaching and assessing clinical competence.* Groningen: Boekwerk Publications, 1990.
15. Rethans J-J, Leeuwen VY, Drop R, *et al.* Performance and competence: two different constructs in the assessment of quality of medical care. *Fam Pract* 1990; **7:** 168-174.



**Address for correspondence**

Dr J-J Rethans, Department of General Practice, University of Limburg, Postbus 616, 6200 MD Maastricht, The Netherlands.

## Chapter 8: Does competence of physicians predict their performance?

Submitted for publication

**Does competence of physicians predict their performance?**
A direct comparison between an examination setting and actual practice using undetected standardized (simulated) patients

**Jan-Joost Rethans, Ferd Sturmans, Riet Drop, Cees van der Vleuten, Pie Hobus**

**Jan-Joost Rethans**, MD, Centre for Quality Assurance in Research in General Practice, Universities of Nijmegen and Maastricht, The Netherlands.
**Ferd Sturmans**, Professor, Department of Epidemiology, University of Limburg, Maastricht, The Netherlands.
**Riet Drop**, Professor, Department of Medical Sociology, University of Limburg, Maastricht, The Netherlands.
**Cees van der Vleuten**, PhD, Department of Educational Development and Research, University of Limburg, Maastricht, The Netherlands.
**Pie Hobus**, MD, Until 1991 Department of Educational Development and Research, University of Limburg, Maastricht, The Netherlands. Since 1991, Family Physician in Wijnandsrade, The Netherlands.

## Summary

Current licensure systems for medical students/doctors are based upon the assumption that they predict actual practice. However, no hard evidence exists for this. Furthermore, it has been shown that doctors in actual practice perform below expert criteria, which led to the question whether they could perform better.

This study investigated 1) the difference and 2) the relationship between performance (what a physician actually does in his daily practice) and competence (what a physician is capable of doing).

39 Family physicians were consulted by four, incognito, standardized patients, portraying four different cases, during normal surgery hours. Later 34 of the 39 doctors participated in a controlled practice test at the Medical School, for which they were asked to act to the best of their ability. Here they were confronted with exactly the same standardized cases (different patients) as they had seen earlier in real practice. Based on the reports of the standardized patients, scores were assigned to the physicians.

Results show that mean competence score of the physicians was substantially higher than their mean performance score. The Pearson correlation between the competence score and the performance score of the participating physicians was negligable. However, when efficiency and consultation time of the consultations were taken into account, moderate to high correlations were found.

It is concluded that performance and competence should be considered as distinct constructs. Assessment of competence under examination circumstances can only have predictive value for performance in actual practice when factors such as efficiency and consultation time are taken into account. Below standard performance of physicians does not necessarily reflect a lack of competence.

## Introduction

Senior and Lloyd distinguish between competence and performance of physicians.[1,2] They define competence as 'what a physician is capable of doing' and performance as 'what a physician actually does in his day-to-day practice'.

In their effort to guarantee quality of medical care, medical schools and other licensing bodies have set up examination systems to decide which students will or will not qualify as practicing physicians.

For this purpose, many countries have installed national examination bodies, whose role it is to develop, spread and apply methods of examination to assess the competence of medical students. Competence consists of knowledge, skills and attitudes.[3] Assessment of competence therefore requires several measurement instruments, each representing different aspects of the construct 'competence'. The choice of a particular assessment method should be based on research on the reliability and validity of the methods.[4] The use of standardized patients in exams (representing the most important aspect of medicine: a consultation with a patient) has been shown to be the most direct method, with a high reliability and high validity.[5-7]

Licensure examinations typically assess competence, whereas assessment of actual practice refers to performance. The assumption behind lisensure examinations is that competence predicts performance: passing examination predicts quality of care and performance in actual practice.

Though intuition tells that this should be the case, surprisingly no evidence exists that this assumption is true. In most studies competence and performance tend to be measured by different methods, or implicitly used concepts are not mentioned.[8] An example of the latter is the use of chart audit as a competence instrument, where in fact chart audit reflects performance. The lack of evidence that competence reflects performance therefore requires the study of the relationship between competence and performance in a sound methodological manner.

The relationship between competence and performance is also important for another reason. Several studies in actual practice show that doctors perform below standards[6, 9-13], whether standards are set by experts[9 - 13] or by the participating doctors themselves.[6] Below standard performance appears to be a consistent finding. It has been suggested that one of the main reasons of this finding is that doctors perform below standard because they lack competence: they simply do not know how to act correctly.[3]

The purpose of the present study was to investigate whether indeed competence and performance as defined by Senior and Lloyd are related.[8,9] For reasons of reliability and validity the method of standardized patients was chosen as an instrument.[5-7] For the performance assessment, unrecognisable standardized patients were introduced into family physicians' normal surgery hours. Recent studies demonstrate that standardized patients may be introduced into practitioners' offices, with negligible detection

rates.[6,9,10] For the competence assessment, standardized patients were used in a controlled examination setting.

It was hypothesized that, as previous studies suggest, achievement of the participating physicians in the competence situation would be higher than in the performance situation.[3,6] On the other hand, it was expected that achievements in both situations would be correlated.

## Subjects and Methods

The study was divided into a performance and a competence part, which took place consecutively.

**Performance part**: From 24 nationally accepted and published Dutch primary care standards, eight medical problems were identified which were all common in general practice, presented a diagnostic challenge and could be presented by a standardized patient.[14] The standards describe obligatory actions (considered to be necessary) and intermediate actions (not essential but not harmful either). Any other actions are considered superfluous. The eight problems were used to construct roles for standardized patients. A panel of three family physicians independently ranked these roles with respect to face validity. The four cases which ranked highest and on which there was agreement among the panel were chosen for our study. The complaints used were 'tension headache', 'acute diarrhea', 'pain in the shoulder' and a 'checkup for a diabetic type II patient'.[14] Diagram 1. represents an example of one of the standards used.

Four groups of three standardized patients (six women/six men) were selected. The three standardized patients in each group had the same gender and approximately the same age. The patients were trained to present a complaint in a standardized manner and to score history-taking, physical and laboratory examination, instructions given to the patient, treatment, and follow-up using the aforementioned standards. To assess the reliability and consistency of scoring among the standardized patients standard procedures were used, which have been carried out successfully before.[9,10] In short, the report of the standardized patient about a consultation with a physician was compared with the report by a panel about the same consultation. The obtained reliability and consistency agreement scores all ranged from 0.8 - 1.0 (Kappa, maximum level is 1.0). All standardized patients signed a written consent to keep all medical and personal information about the general practitioners in this project strictly for research purposes.

In January 1988 all 442 general practitioners working in the province of our university were informed about the study and asked to give their written acceptance of standardized patients into their practices for a period lasting three years and of later participation in the competence part. The doctors were not told how often or when they would be visited, nor the content of consultations, but would be informed when and by whom they had been

visited at the end of the project. They were asked to report every patient whom they thought they detected as a standardized patient. Four months before each of the planned visits, the standardized patients were enlisted in the practices of the participating doctors using techniques reported earlier.[15,16] The standardized patients made visits during a four months period, at least 12 months after the doctors agreed to participate.

**Competence part**: Five months after the visits of the standardized patients, the participating doctors were invited to the Medical School. They were installed in rooms which had been fitted out like consultation rooms of family physicians. They were then instructed, orally and in writing, to perform to the best of their abilities, free of time pressure, in their subsequent contacts with a number of standardized patients. The doctors were told that the quality (not the quantity) of their consultations was going to be assessed. The doctors were not told the number or content of cases. All consultations were audiotaped and videotaped. The same standardized patients were used and tested again for their reliability and consistency (kappa-scores ranged from .78 to .94). The case histories used were the same as in the performance part, but no doctor met the same standardized patient as in the performance part.

Several variables were chosen to measure different aspects of competence and performance.

First, performance and competence scores for each doctor were calculated by counting the number of obligatory, intermediate and superfluous actions, leading to 'obligatory, intermediate and superfluous scores'. In addition, a 'total score' was calculated by summing all obligatory, intermediate and superfluous actions.

Second, since the obligatory actions (and hence their score) were regarded as the most essential part of a consultation, the ratio between the obligatory and total scores was calculated for each complaint and across the four complaints. This ratio was defined as 'efficiency score', since a high score on this variable reflects physicians with relatively more obligatory actions compared to additional actions, suggesting an efficient consultation.[10]

Third, recent cognitive literature on medical problem solving suggests that the time a physician uses to solve a medical problem reflects an important aspect of expertise.[17-19] It has, for example, been shown that the longer it takes a doctor to state a diagnosis, the more likely it is that his diagnosis is wrong.[20,21] Furthermore, within a certain time span expert physicians show better and more adequate processing of relevant patient information than less experienced colleagues, suggesting that time combined with efficiency is important.[17,22] In concordance with this line of reasoning, two other variables were calculated. First, for each case and across the four cases, the duration of the consultations (in minutes) was measured; the 'time score'. Second, the above-mentioned 'efficiency score' was divided by this time score. This new score reflects the level of efficiency per unit of time, and was hence named 'efficiency-time score'. High efficiency-time scores

**Diagram 1. Standard of the Headache case**

**History**

1. Obligatory. Time aspects: how long have complaints existed; at what times of the day; how often?
2. Obligatory. Nature of the pain.
3. Obligatory. Location of the pain, radiation, presence of prodomes; progression of complaints.
4. Obligatory. Associated phenomena (such as light phobia, nausea, fever, dental and neck complaints,etc.).
5. Obligatory. Relation with psychosocial circumstances.
6. Intermediate. What is the reaction to the pain.
7. Intermediate. Self therapy.
8. Intermediate. Intoxications (smoking, carbon monoxide).
9. Obligatory. Migraine history (familial history, progression, frequency, connection with circumstances).
10. Superfluous actions history.

**Physical examination**

11. Intermediate. Blood pressure reading
12. Intermediate. Examine eye sight.
13. Intermediate. Examine eye fundus.
14. Intermediate. Cervical spine.
15. Intermediate. Neurological examination.
16. Intermediate. Sinuses: percussion and pressure pain, transillumination.
17. Superfluous actions physical examination.

18. Superfluous actions Laboratory

**Guidance and advice**

19. Obligatory. Explain cause of complaints.
20. Obligatory. Discuss prognosis.
21. Obligatory. Explain relationship between complaints and tension.
22. A. Obligatory. In case of therapy: explain expected effect.
    B. Obligatory. In case of no therapy: explain why no therapy is prescribed.
23. Intermediate. Relaxation exercises, yoga (brochure).
24. Obligatory. Discuss connection with life style.
25. Superfluous actions guidance and advice.

**Therapy**

26. Intermediate. Simple analgetics.
27. Obligatory. Discuss possible background to headache.
28. Intermediate. Benzodiazepines.
29. Superfluous actions therapy.

**Return visit**

30. Obligatory. Indicate whether or not a return visit is necessary, depending on possible increase of complaints and prognosis.
31. Superfluous actions return visit.

differ from low ones in that these consultations contain relatively more obligatory actions per unit of time.

First the Wilcoxon signed rank test (paired design) was used to look for differences in the doctors' scores in the competence and performance parts. Secondly, observed correlations (Pearson product moment correlations) were calculated between doctors' scores on the two formats (disattenuated "true" correlations were not calculated, because equal content of cases were used in both formats).

## Results

Of the 442 doctors asked to participate 137 (31%) agreed to be visited, of whom 131 also agreed to take part in the competence part. 39 of them were selected and visited; three of them as pilot practices. For financial reasons, the main selection criterion for including a general practitioner was the distance between his/her practice and the university (less than 30 kilometres.) After all visits had taken place, 36 doctors were asked to take part in the competence part, and 34 (94%) agreed to do so. Personal and practice characteristics of the participating doctors did not deviate from national data (mean years of practice-experience of these 34 doctors was 12 (range 1-28) and 15 (44 %) of them were solitary working). In none of the 156 visits in actual practice was a standardized patient detected and none of the doctors was able to recall these cases, before ór after the competence phase.

Table 1 shows the obligatory, intermediate, superfluous, total and time scores, as calculated from the actions performed in the performance and competence parts. The obligatory and intermediate scores have a maximum number of actions per standard, so these scores can also be calculated as percentages of a particular standard. These percentages are also shown in Table 1, under heading "% of standard". Since the number of superfluous actions is potentially infinite, no meaningful percentages can be calculated for them. From Table 1 it follows that there was a significant difference between competence and performance for each of the variables, with competence scores being consistently higher than performance scores. The mean total score for competence across four complaints, for example, showed an increase of 49.5% compared with the same performance score. There was also a difference in the obligatory performance scores between the first three complaints and the diabetic case with respect to adherence to the standards, respectively 69, 58, 65% and 34%.

Table 2 shows the mean efficiency scores across four complaints and for individual complaints. This table shows again that there was a significant difference between competence and performance across four cases and for three of the separate cases, but now in favour of performance. It seems that when it came to efficiency, doctors did better in actual practice than in a test. The same effect is also found in Table 3, which shows the mean efficiency-

**Table 1.** Mean obligatory, intermediate, superfluous, total and time scores (range and standard deviation) and percentages of the standard for obligatory and intermediate scores across four cases and for each individual case as reported by standardized patients visiting 34 general practitioners, for, respectively, the performance and competence settings.

| | Mean | Range | Standard deviation | % of standard |
|---|---|---|---|---|
| **Across four cases** | | | | |
| Obligatory score (Max= 68) | | | | |
| Performance | 37.08 **** | 25-51 | 6.35 | 55 |
| Competence | 49.05 | 38-58 | 5.27 | 72 |
| Intermediate score (Max= 29) | | | | |
| Performance | 8.52 **** | 3-14 | 2.59 | 29 |
| Competence | 12.44 | 9-16 | 2.19 | 42 |
| Superfluous score | | | | |
| Performance | 9.11 **** | 2-19 | 3.99 | |
| Competence | 20.32 | 10-35 | 6.68 | |
| Total score | | | | |
| Performance | 54.70 **** | 33-83 | 10.05 | |
| Competence | 81.82 | 61-102 | 10.97 | |
| Time score | | | | |
| Performance | 38'40" **** | 19-69 | 12'05" | |
| Competence | 55'38" | 27-100 | 16'23" | |
| **Headache Case** | | | | |
| Obligatory score (Max= 13) | | | | |
| Performance | 9.02 *** | 5-12 | 1.62 | 69 |
| Competence | 10.64 | 7-13 | 1.25 | 82 |
| Intermediate score (Max= 12) | | | | |
| Performance | 4.08 *** | 0-7 | 1.65 | 34 |
| Competence | 5.58 | 3-9 | 1.63 | 47 |
| Superfluous score | | | | |
| Performance | 2.70 *** | 0-7 | 1.91 | |
| Competence | 5.23 | 1-13 | 2.73 | |
| Total score | | | | |
| Performance | 15.82 **** | 11-21 | 2.79 | |
| Competence | 21.47 | 13-29 | 4.17 | |
| Time score | | | | |
| Performance | 11'33" **** | 4-19 | 3'45" | |
| Competence | 17'07" | 7-28 | 5'01" | |

| | Mean | Range | Standard deviation | % of standard |
|---|---|---|---|---|
| **Diarrhea Case** | | | | |
| Obligatory score (Max= 15) | | | | |
| Performance | 8.67 **** | 5-14 | 2.4 | 58 |
| Competence | 12.08 | 9-15 | 1.5 | 81 |
| Intermediate score (Max= 5) | | | | |
| Performance | 2.50 * | 1-4 | 1.05 | 50 |
| Competence | 2.94 | 1-5 | 0.91 | 59 |
| Superfluous score | | | | |
| Performance | 1.88 ** | 0-8 | 1.88 | |
| Competence | 3.17 | 0-8 | 1.78 | |
| Total score | | | | |
| Performance | 13.05 **** | 6-23 | 4.03 | |
| Competence | 18.20 | 13-23 | 2.38 | |
| Time score | | | | |
| Performance | 6'47" **** | 3-12 | 2'36" | |
| Competence | 9'42" | 4-18 | 3'32" | |
| **Shoulder pain Case** | | | | |
| Obligatory score (Max= 19) | | | | |
| Performance | 12.29 ** | 5-16 | 3.13 | 65 |
| Competence | 14.23 | 9-17 | 1.89 | 75 |
| Intermediate score (Max= 6) | | | | |
| Performance | 1.76 ** | 0-3 | 0.92 | 29 |
| Competence | 2.67 | 1-5 | 1.17 | 44 |
| Superfluous score | | | | |
| Performance | 2.50 **** | 0-7 | 1.60 | |
| Competence | 5.08 | 1-9 | 1.91 | |
| Total score | | | | |
| Performance | 16.55 **** | 6-24 | 4.26 | |
| Competence | 22.00 | 14-28 | 3.42 | |
| Time score | | | | |
| Performance | 8'00" **** | 3-13 | 2'56" | |
| Competence | 12'08" | 5-26 | 4'28" | |
| **Diabetic Case** | | | | |
| Obligatory score (Max= 21) | | | | |
| Performance | 7.08 **** | 2-14 | 2.81 | 34 |
| Competence | 12.08 | 6-19 | 3.03 | 58 |
| Intermediate score (Max= 6) | | | | |
| Performance | 0.17 **** | 0-7 | 0.45 | 0 |
| Competence | 1.23 | 0-3 | 0.78 | 21 |
| Superfluous score | | | | |
| Performance | 2.00 *** | 0-7 | 1.79 | |
| Competence | 6.82 | 0-24 | 5.72 | |
| Total score | | | | |
| Performance | 9.26 **** | 2-18 | 4.04 | |
| Competence | 20.14 | 8-45 | 7.87 | |
| Time score | | | | |
| Performance | 12'19" ** | 4-36 | 6'55" | |
| Competence | 16'40" | 7-28 | 5'53" | |

**** $p < 0.0000$; *** $p < 0.0005$; ** $p < 0.005$; * $p < 0.05$; Wilcoxon signed rank test (paired design)

**Table 2.** Mean efficiency score (range and standard deviation) across four cases and for each individual case as reported by standardized patients visiting 34 general practitioners, for, respectively, the performance and competence settings.

| | Mean | Range | SD |
|---|---|---|---|
| **Across four complaints** | | | |
| Performance | .68 **** | .57 - .84 | .06 |
| Competence | .60 | .50 - .68 | .04 |
| **Headache complaint** | | | |
| Performance | .58 ** | .35 - .91 | .12 |
| Competence | .50 | .34 - .73 | .09 |
| **Diarrhea complaint** | | | |
| Performance | .67 | .50 - .91 | .10 |
| Competence | .66 | .34 - .84 | .07 |
| **Shoulder complaint** | | | |
| Performance | .74 **** | .53 - .91 | .08 |
| Competence | .65 | .42 - .85 | .08 |
| **Diabetes complaint** | | | |
| Performance | .79 ** | .45 - 1.0 | .14 |
| Competence | .63 | .38 - .90 | .13 |

**** $p < 0.0000$; ** $p < 0.005$; Wilcoxon signed rank test (paired design).

**Table 3.** Mean efficiency-time score (range and standard deviation) across four cases and for each individual cases as reported by standardized patients visiting 34 general practitioners, for, respectively, the performance and competence settings.

| | **Mean** | **Range** | **SD** |
|---|---|---|---|
| **Across four complaints** | | | |
| Performance | .019 **** | .009 - .036 | .006 |
| Competence | .011 | .005 - .023 | .004 |
| **Headache complaint** | | | |
| Performance | .058 **** | .027 - .211 | .034 |
| Competence | .034 | .092 - .104 | .018 |
| **Diarrhea complaint** | | | |
| Performance | .121 **** | .050 - .277 | .062 |
| Competence | .080 | .036 - .176 | .035 |
| **Shoulder complaint** | | | |
| Performance | .108 **** | .053 - .277 | .049 |
| Competence | .060 | .026 - .150 | .024 |
| **Diabetes complaint** | | | |
| Performance | .084 **** | .020 - .250 | .049 |
| Competence | .044 | .016 - .093 | .020 |

**** $p < 0.0000$; *** $p < 0.0005$; Wilcoxon signed rank test (paired design).

time scores for the competence and performance settings. Across four cases, for example, the mean efficiency-time score in the performance setting was 65% higher than its equivalent competence score.

Table 4 shows the correlations between competence and performance for each of the variables used. The table shows several surprising outcomes. The correlations of .00 and -.04 for, respectively, the obligatory score and the total score across four
complaints, suggests a disappointingly low (virtually no) correlation between competence and performance. The same conclusion can be drawn for the efficiency score. Surprisingly, this score shows correlations which do not deviate much from the intermediate and superfluous scores. In contrast, the time and efficiency-time scores across four cases seem to indicate a more substantial correlation between competence and performance. The individual cases in Table 4 show the same interesting pattern: very low correlations for the obligatory and total scores, but high values for the time and efficiency-time scores.

However, as in Table 1, the diabetic case of Table 4 again shows a pattern that is different from the other complaints, with many negative values, in contrast to the other cases. For this reason we also calculated what the correlation between competence and performance would be without the diabetic case. The results are also included in Table 4. These values (without diabetes) show again that the obligatory and total scores yield a low correlation. The efficiency-time score, however, yielded high correlations, up to .77, between competence and performance.

**Table 4.** Pearson product moment correlation between competence and performance for the obligatory, intermediate, superfluous, total, time, efficiency and efficiency-time scores across four cases, across three cases and for each individual case with 34 physicians.

| | Across four complaints | Head-ache | Diarrhea | Shoulder | Diabetes | Across Three complaints (minus diabetes) |
|---|---|---|---|---|---|---|
| Obligatory score | .00 | .07 | -.11 | .20 | .15 | .21 |
| Intermediate score | .26 | .33 | .25 | .07 | -.04 | .32 |
| Superfluous score | -.05 | .24 | .25 | .10 | -.25 | .25 |
| Total score | -.04 | .23 | .24 | .07 | -.07 | .29 |
| Time score | .49 ** | .48 ** | .67 *** | .47 ** | .15 | .61 *** |
| Efficiency score | -.08 | .34 * | .32 | .28 | -.30 | .35 ** |
| Efficiency-time score | .45 ** | .72 *** | .59 *** | .62 *** | .00 | .77*** |

* $p < 0.05$; ** $< 0.01$; *** $p < 0.001$

## Discussion

This is the first time a study has compared actual performance of doctors in their practices with data collected in an examination setting, both assessed with the direct method of standardized patients. Clearly, the study has some limitations.

First, competence has been reduced to its medical-technical aspect, which means that physicians' attitudes and knowledge were not studied.

Second, this study used only four cases per doctor, which is too few to allow generalization beyond the specific sample of cases used.[7]

Third, 31% of invited doctors responded positively. The question arises whether these doctors represent a selected group with higher standards of care than non-participants, although personal and practice characteristics of the participating doctors did not deviate from national data. If it is assumed that the doctors under study are the better ones, then it remains unclear whether the competence-performance difference of the non-participants is likely to have been larger or smaller.

Fourth, there is the validity question of the standards: would different standards have produced different results? As was explained, the standards were developed for use in actual practice. As the designers were well aware of the risk of setting expert criteria, the construction procedure consisted of several rounds of comments including a pilot phase in practice.[16] Therefore, at the start of this study there was no reason to doubt the validity of these standards.

With these limitations in mind, several important results emerge from this study.

First, from the large, significant differences between the competence and performance scores for all the variables used, it can be concluded that there is a substantial difference between competence and performance of physicians.

Second, the direction of the differences in scores between the two formats is very interesting and challenging. The use of only quantitative data (obligatory, intermediate, superfluous and total score) shows that doctors did more in the competence than in the performance situation. However, using qualitative data (efficiency and efficiency-time score), it was in actual practice that doctors did better.

Third, a major finding of this study is that it appears to be crucial to choose the correct variable to measure the correlation between competence and performance. This is shown by the differences in level between, on the one hand the obligatory, total and effiency correlations and, on the other hand the time and efficiency-time correlations. It is difficult to explain why the level of the efficiency score correlation is not higher than those of the obligatory, intermediate and superfluous scores. Perhaps this reflects that scores, without the time-factor, are bad predictors of performance. The high time correlation seems to indicate that time is also an important factor with respect to

expertise. As regards the differences between the first three cases and the diabetic case, the participating doctors were asked about the reasons for this difference. They responded that the diabetic case, in contrast with the others, reflected a chronic disease, for which in general practice several consultations would normally be used in order to gather all necessary patient data. The type of problems, therefore, remains an important issue in real practice as well.

We believe the implications of the study are in two areas, viz. the examination-setting or licensing area and the area of setting standards for the quality of care.

First, since the doctor-patient consultation is the final purpose of medical education, assessment of students' or physicians' capacities in this contact is crucial. This study shows that, if qualitative data are used in combination with the time that examinees need to perform a test, competence will indeed be predictive of performance. However, examinations are generally restricted to quantitative data. This study shows that these data are poor predictors of performance, casting serious doubts about the validity of current examination systems. It seems crucial to make use of proper variables in order to be able to predict performance from competence situations. Second, with respect to standards of care it can be concluded that this study shows that doctors perform below standard in actual practice, but that they are able to perform significantly better when asked to do so. It can be concluded that the physicians are more competent than their performance in practice shows. On the other hand in the competence test they also perform below the standards. Taking into account other experiments with internal or external standards, it might be asked whether the procedures of setting standards are valid.[6, 9-13] Even when these procedures contain pilot-practice phases and even when doctors are asked to formulate standards themselves, results show that doctors perform below standards. The finding in this study that the participating doctors had good reasons not to adhere to the diabetic standard, shows that one has to be cautious about interpreting the scores of physicians on standards. The finding that doctors act more efficiently in real practice than in a test, also means that one should be cautious in simply concluding that doctors just have to do better, in relation to the standards, than in real practice. It has been shown before, also with standardized patients, that doctors in actual practice show efficient performance.[10]

The participating physicians were left free of time pressure in the competence part of our study. The reason for doing so was that we wanted to investigate whether doctors would indeed be able to perform more actions in the competence situation than in actual practice. It would be interesting now to investigate what would happen if doctors were to perceive time pressure in the competence setting as well.

It seems necessary to reconsider carefully the procedures for setting standards. It might be advisable to start with an assessment of actual practice in a valid way (by observing how doctors really perform). In this process standardized patients can play an important role. Since the standardized

patients were highly satisfied in nearly all consultations, even though the doctors varied in their level of scores, it is suggested that the number of actions performed alone can never be the sole base for assessing a physician's competence.

From this study it can be concluded that there is a difference between competence and performance of physicians. Taking qualitative data into account, competence is a predictor of performance. This should be taken into consideration in examination systems and in decisions about the way doctors have to be assessed for (re)certification reasons.

Quality of care in actual practice should not be assessed on the basis of standards alone. Other aspects of actual practice, such as patient satisfaction, consultation time and outcome of consults, should also be taken into consideration.

## Acknowledgments

We are grateful to the 39 participating family physicians in Limburg, to Caroliene Janssens, Marike Laning, Marijke Verdonk, Bea de Grootte, Martha Lucassen, Juul Kerbusch, Peter Kramer, Pieter Ramler, Pierre Bastings, Leon Heuts, Wil Macco, Trudie Seegers, the sickfunds VGZ in Maastricht and LIASS in Heerlen and Sittard, and Robin Hull of the General Practitioners Writers Association for their efforts in this study.

## References

1. Senior JR. Towards the measurement of competence of medicine. Philadelphia: National Board of Medical Examiners 1976.
2. Lloyd JS. Definitions of competence in specialties of medicine. Chicago: American Board of Medical specialties 1979.
3. Neufeld VR , Norman GR. Assessing clinical competence. Springer Publishing Company, New York, 1985.
4. Maguire P. Assessing clinical competence. Br Med J 1989; 298: 4-5.
5. Stillman PL, Swanson DB. Ensuring the clinical competence of medical graduates through standardized patients. Arch Intern Med 1987; 147: 1049-52.
6. Norman GR, Neufeld VR, Walsh A, Woodward CA, McConvey G. Measuring physicians' performances by using simulated patients. J Med Educ 1985; 60: 925-34.
7. Vleuten CPM van der, Swanson DB. Assesment of clinical skills with standardized patients: state of the art. Teaching and Learning in Medicine 1990; 2: 58-76.
8. Rethans JJ, Leeuwen v Y, Drop R, Sturmans F, Vleuten vd C. Performance and competence: two different constructs in the assessment of quality of medical care. Fam Pract 1990; 7: 168-74.
9. McClure CL, Gall EP, Meredith KE, Gooden MA, Boyer JT. Assessing clinical judgment with standardised patients. J Fam Prac 1985; 20: 457-64.
10. Rethans JJE, Van Boven CPA. Simulated patients in general practice: a different look at the consultation. Br Med J 1987; 294: 809-12.
11. Norman GR, Tugwell P, Feightner JW. A comparison of resident performance on real and simulated patients. J Med Educ 1982; 57: 708-15.
12. Lomas J, Anderson GM, Domnick-Pierre K, et al. Do practice guidelines guide practice? N Engl J Med 1989; 321:1306-11.

13. Grol R. Kwaliteitsbewaking in de huisartsgeneeskunde. Krips Repro Meppel, 1986. Thesis. (Summary in english)
14. Grol R, Mesker P, Schellevis F (eds.). Peer review in general practice. Nijmegen University Department of General Practice. 1988: 87-9; 97-8; 107-9; 117-9.
15. Rethans JJ, Drop R, Sturmans F, Vleuten vd C. A method for introducing standardized (simulated) patients into general practice consultations? Br J Gen Pract 1991; 41: 94-6.
16. Rethans JJ, Sturmans F, Drop R, Vleuten vd C. Assessment of the performance of general practitioners by the use of standardized (simulated) patients. Br J Gen Pract. 1991; 41: 97-99.
17. Schmidt HG, Norman GR, Boshuijzen HPA. A cognitive perspective on medical expertise: Theory and implications. Acad Med 1990; 65 : 611-21.
18. Schmidt HG, Boshuijzen HPA, Hobus PPM. Transitory stages in the development of medical expertise: the intermediate effect in clinical case representation studies. In Proc. 10th Conf Cognitive Science Soc. Hillsdale, New Jersey: Erlbaum; 1988: 139-45.
19. Norman G, Allery l, Berkson L et al. Research in the psychology of clinical reasoning: implications for assessment. In: Jolly B (ed). New directions in the assessment of clinical competence. Proc Cambridge Conf IV, Madingley Hall, Cambridge. In press.
20. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision making in pulmonary nodule detection. Invest. Radiol. 1987; 13: 175-81.
21. Norman GR, Brooks LR, Allen SW, Rosenthal D. The development of expertise in dermatology. Arch Dermatol 1989; 125: 175-81.
22. Hobus PPM, Schmidt HG, Boshuizen HPA, Patel VC. Contextual factors in the activation of first hypotheses: expert-novice differences. Med Educ 1987; 21: 471-76.

# Chapter 9: Conclusions and recommendations for research

In the introductory chapter it was stated that the papers in this book had two main aims.

The first of these was to investigate whether it was possible for a set of standardized (or simulated) patients to visit general practitioners during normal surgery hours without being detected. The second aim was to investigate 1) the difference and 2) the relationship between the competence and performance of general practitioners, with performance defined as "what a doctor does in his day-to-day practice" and competence as "what a doctor is capable of doing".[1,2]

In this chapter the main conclusions of the projects will be discussed. Some advice for further research in the field of standardized patients and of competence and performance will be given.

## The methodology of standardized patients.

In the first of the two real practice experiments described here, only two out of 48 visits by standardized patients have been detected, and none in the 156 consultations of the second project. Two patients in the second project who were reported to be standardized patients turned out to be genuine. Together with the results of the assessment of the consistency of the standardized patients (kappa 0.78 to 1.0), these findings show the feasibility of this method in a health care system, even where doctors have fixed patient lists.

The main reason for this success probably lies in the detailed procedures used for both the training and for the actual visits by the standardized patients. In discussions with other researchers involved in the training of standardized patients at the Fourth Ottawa Conference on Assessing Clinical Competence (July 1990), it became obvious that the length of the training (role playing and reliability training) of our patients (14 hours) was rather long.[3] The other researchers mentioned having used about half the training time we used in our projects. However, they did not have experience with the use of standardized patients in real practice and it was for this real practice purpose that we in our project were more afraid of using too short a period of training than one which was too long. It is difficult to estimate whether the procedures used in our projects were too detailed; in the words of one of the participating physicians: "How much of your preparation has been superfluous?". Some anecdotal reports by our patients seemed to indicate that nothing was superfluous when it came to preventing the detection of standardized patients. For example, as explained in chapter 6, the selection of home addresses for the standardized patients was an important feature in the preparation phase for the actual visits. If it was impossible to find a credible address by searching in the computer of another doctor with a practice in the same city, and if it was also impossible to "borrow" adresses from friends or relatives of the researchers, a risk was taken by selecting a non-existing house number in a known street or by

selecting a random number in an apartmentbuilding. One standardized patient, who was supposed to live at such a random number in an apartmentbuilding, reported that at the start of the consultation the general practitioner he was going to visit showed him a medical file with the same surname and address as the one he had just been given, and asked: "Is she your sister?". So, accidentally we had given the standardized patient both the same name and the same address as an genuine patient. In this case it was enough for the standardized patient to present the medical assurance card to convince the doctor that the (real) address on the existing medical file was likely to be false.

Although during a large number of visits such details as 'being familiar with the neighbourhood of the doctors visited' was not a point which was discussed during the actual contacts, these details were crucial during some other visits. They also helped the patients to feel secure during the visits.

In general, it can therefore be said that as a means of preventing detection of any standardized patient, nothing in the preparation phases was superfluous. The question is however whether it is necessary that not a single case is detected. To answer this question a series of experiments would be necessary.

At the start of these projects the majority of general practitioners in the region of the university were unfamiliar with the method of standardized patients, since it had never been introduced in actual practice. Given this fact, we estimate that the figure of 31 % of invited doctors who agreed to participate in this project is a high one.

Although the personal characteristics of the participating doctors and their practices did not deviate from the national characteristics, we cannot exclude the possibility that the participating doctors represent the "cream" of all doctors.

Several months after the visits had taken place we informed the participating doctors about the nature and content of the visits. All of them reacted positively to the visits, that is to say that they had no objection to the fact that details of the medical part of their contacts with the patients were to be reported. We were told that the guarantee of anonymity given to the participating doctors helped a great deal in creating this positive feeling. We think that in future, with the results of the present book at hand, the percentage of doctors who will be willing to participate will increase.

What are the prospects for the method of standardized patients in real practice after the experiments described here? In other words, will the method be used more often, and for what purposes?

Firstly, the experiments have led to a discussion about the way in which audit will have to be organized in practice. This discussion was reflected by several letters to medical journals and other media.[4-8] The main issue in the debate about the prospects for this method seems to be whether the method should be used only if doctors give prior permission for the introduction of standardized patients. Although this seems a reasonable demand from the

point of view of the doctors and their organizations, it may well lead to a different opinion when viewed from the perspective of the patients. In fact, the use of standardized patients in real practice consitutes audit by consumer and we tend to consider that the consumer's opinion is more important than that of the doctors.

Secondly, our data collected with the standardized patient method is restricted to the medical technical findings during the consultations. However, in a non-systematic way we have also reported, in chapter seven, on the opinions of the standardized patients as regards the caring qualities of the participating doctors. Research with standardized patients for educational and assessment purposes has shown that standardized patients are also capable of reporting in a consistent manner on the social and technical skills of medical students.[9] It would therefore be worthwhile to extend research in real practice with standardized patients, reporting also on the social and attitudinal skills of physicians.

Thirdly, with regard to research on inter-doctor variation the method of standardized patients will be potentially the best method, since by using standardized patients researchers will be able to present the same stimulus (the standardized patients) to a set of different doctors and be able to compare the difference in performance between doctors. Of course, also in the future, whether the method of standardized patients is the best one to use will still depend on what researchers are looking for in actual practice. For very detailed, for example semantic, analysis of consultations, other methods such as videotaping or audiotaping may be even better.

However, with regard to the assessment of the performance of doctors in their day-to-day care of patients, the method of standardized patients is the best one to use. It is also our experience that this type of research is furthermore very enjoyable for researchers, for standardized patients and for the participating doctors.

## Competence and performance.

Several interesting results have emerged from the research conducted.

Firstly, the findings show that all the doctors participating performed substantial more actions in the competence setting than in the performance setting. This result was consistent for all categories used: obligatory, intermediate and superfluous actions. It is unequivocally clear that doctors perform more of the actions mentioned in the standards in a test setting than in actual practice, something which was a matter of debate before our project started.

Secondly, an interesting finding was also that when it comes to efficiency (with efficiency defined as the ratio between the obligatory actions and the number of total actions per consultation), doctors performed substantially better in real practice than in the test setting. This finding suggests that

doctors may have very good reason for carrying out, or not carrying out, certain actions in practice. For example, this was well demonstrated by the reaction of the participating doctors to their performance in the diabetic case. When the doctors were asked for comments on their low performance (according to the standard used), they argued that when dealing with chronic patients (and complaints) they did not intend to perform all the actions during a single visit. They stated, that working as a general practitioner means that patients are being seen in a longitudinal manner, i.e. during consecutive visits thoughout the years, giving ample opportunities gather data. They argued that the diabetic standard did not take account of this real life fact.

Future research in this area may be focussed on the reasons for the differences found and on the reasons why certain actions were or were not performed. One way to conduct such a study is to introduce standardized patients into general practitioners' practices and to inform the doctors immediately after the consultation, that they have been visited. Then, with the report of the standardized patient at hand, the researchers could ask the doctors, for each of the reported items, for the reason for performing or not performing medical actions. The design of such a study could also be used to study the effects of feedback on real performance of physicians. Feedback in real practice by means of a standardized patient immediately after a consultation is probably very effective, since experience with it for educational purposes is well described.[10,11]

Thirdly, a major finding of this study is that it appears to be crucial to choose the correct variable to assess the association between competence and performance. This is shown by the difference in level of the competence-performance correlations, between on the one hand the obligatory, total and efficiency (competence-performance) correlations and, on the other hand the time and efficiency-time (competence-performance) correlations.
It is difficult to explain why the efficiency score (competence-performance) correlation is not higher than the competence-performance correlations for the obligatory, intermediate and superfluous scores. It suggests that scores without the factor time are bad predictors of performance. The high correlation on the variable "time" seems to indicate that time is an important variable with respect to expertise.

A recent cognitive theory as to the development of medical expertise may clarify this finding.[12] In this paper a new theory for the development of medical expertise is presented: "Contrary to existing views, this theory assumes that expertise is not so much a matter of superior reasoning skills or in-depth knowledge of pathofysiological states as it is based on cognitive structures that describe the features of prototypical or even actual patients. These cognitive structures, referred to as "illness scripts", contain relatively little knowledge about pathofysiological causes of symptoms and complaints but a wealth of clinically relevant information about disease, its consequences, and the context under which illness develops".

The authors describe their theory as a stage theory: during their medical education students/doctors pass through several hiearchical stages, each representing a phase in the ascent to expertise.

The first stage is one of forming so-called "propositional or conceptual networks". When confronted with medical problems and solving them students search for events, concepts or objects which can be related to each other and they then try to make causal networks out of them.

The second stage starts with their being confronted with real patients. Having met the same case or problem before (or cases with same features to those already met), shortcuts will emerge in the student's diagnostic process. His knowledge-in-use will reorganize itself (by shortcuts) so that accessibility and efficient use are assured.

The third stage is one of the emergence of "illness scripts". This may be simplified by imaging that, simultaneously with a further compilation of knowledge (stage 1-2), students begin to pay attention to contextual factors under which diseases emerge. Instead of causal processes (stage 1-2), the different features that characterize the clinical appearance of a disease become more important. Problem solving in routine cases thus becomes a process of "script search, script selection and script verification".

The fourth and final stage is called "storing patient encounters as instance scripts". In its essence this stage is characterized by the finding that experienced doctors solve cases by recognizing a new presentation simply on the basis of its similarity to one previously encountered. The theory would predict that when dealing with medical problems which they have not seen before, doctors switch to stages which are lower in the hierarchical system presented. This means that physicians dealing with a more or less unfamiliar problem try to solve it with knowledge/experience at a more basic level and start reasoning towards the solution of a problem on this basis.

In their arguments for presenting this theory, the authors repeatedly discuss the influence of time on the process of clinical reasoning. They argue that the amount of time a physician needs to solve a medical problem reflects an important aspect or results of expertise, in that less time needed reflects better expertise.[13-15] More experienced physicians make more use of "instance scripts", and this takes less time then problem solving. Furthermore, within a certain time-span, expert physicians compared to less experienced colleagues show better and more adequate processing of relevant patient information, suggesting that time combined with efficiency is important.[13,16]

Our data does not allow accurate measurement of differences between experienced and less experienced doctors, since the range of years of experience of our doctors is too small, but the issue of time should certainly be taken up in new studies. In the competence part of our study we have stressed that the doctors should take their time, but we did not say "please take as much time as you can". Of course the effects of this remark to the doctors can be regarded as a bias: all doctors used more time than in the

practice setting and they thus performed more actions. As we have said earlier in this discussion, a major aim of the study was to see if doctors could perform better. We therefore intended to give the doctors as much time as they desired. Suppose we had put the doctors under stress in the competence part by saying they would be rewarded by taking a very short time, or by saying that they would be punished for using too long a time, or by saying that they were supposed to treat 40 patients in two hours. If the result of such a study had been that doctors performed fewer actions in the competence part than in the performance part for all variables used, the design could be criticized for not giving the doctors enough leeway to spread their talents.

With the results of the current study at hand it is advisable to conduct the same study as we did, but now with two random groups: one with and one without "time-stress" in the competence part. In this way the same variables (and their competence performance correlations) as we used could be compared under two different conditions.

On the basis of our results other experiments with the factor time might also be worthwhile. We would like to suggest studying what will happen with the results of medical students in examination settings when these results are corrected for the amount of time students take to solve the problem. So far, this has not been usual in current examination systems.[17] When two students score the same number of points on the Maastricht Progress Test and one of them took four hours and the other only one hour to solve the problems, the results are the same: the same number of points. If tests are supposed to differentiate between levels of medical expertise, then possibly time could be taken into consideration in the outcome.

There is one important remark, or rather restriction, to be made with regard to the conclusions about the factor time in our study. We have not measured time per unit or part of a consultation. For example, the stage-theory deals with time from a cognitive perspective, that is to say that time is measured only with respect to stating a diagnosis or to solving a problem. Besides the cognitive part of a consultation our measurement of time also contains other parts, such as time to shake hands, time to show the patient the way to the pharmacy, time to explain the diagnosis, time to help (un)dress the patient, time to fill in a lab form etc. This implicates that although time seems to be an important factor, more precise measurement of time (for example per unit of consulatation) may well show that in some respects time may be less important. This shows that future studies with respect to the impact or result of time need to specify which component of the consultation the time is needed for.

## Standards of care

The studies reported in this book show that doctors in real practice perform below standards of care. The mean numbers of actions performed by the doctors in real practice were, for the obligatory category according to the standards, 69, 58, 65 and 34% for respectively the headache, diarrhoea, shoulder and diabetes cases.
Since below-standard performance has been demonstrated before, this may not be seen as an innovative finding. It is therefore more interesting to compare these performance values with their equivalent competence values, which are 82, 81, 75 and 58 % respectively. These values show without doubt that doctors are able to perform better in a test situation than in real practice.

Things get complicated, however, with the finding that with regard to efficiency doctors do better in the real practice situation than in the test situation.

What does this result mean for the current policy of standard setting in primary care? We think that the results show that one has to be careful when drawing conclusions about doctors' performance, if these conclusions are only based on scores on standards, reflecting competence assessments. The results also show that one has to take account of what actually happens in general practice. We have asked the participating doctors, in a not-systematic manner, about their explanation for the large difference in performance scores between the diabetic case and the other three cases and we have discussed this already (section "competence and performance"). Their main criticism was that the diabetic standard did not take into account the longitudinal aspects of consultations in general practice: patients will be seen at consecutive points in their life and during illness. This longitudinal aspect of general practice is not only very important in chronic diseases, it is also important in more or less acute diseases. The knowledge a specific general physician has about a specific patient with regard to such factors as age, sex, family, use of medication, number of visits per year etc (factors often known as "contextual factors") has been shown to be very important for the diagnostic aspect of consultations.[18] We think that the finding in our study that doctors show more efficient behaviour in real practice (and in less time) than in the competence part is a reflection of this phenomenon. If standards in general practice are to be effective with regard to an increase in the quality of general practice (as is the target of the policy makers who produce standards), then standards should at least cover both the longitudinal and the contextual aspects of general practitioners' work. However hardly any of the current standards seems to cover these issues.

Does this mean that standards are useless, a waste of time? No, we do not think so. They are certainly valuable, but until now mainly for educational reasons. The process of setting standards in general practice is in its infancy, and the Dutch College of General Practitioners (Nederlands Huisartsen

Genootschap) is well ahead of other national colleges with respect to the process of standard setting. The Dutch College should however be aware of the fact that the process by which standards are currently being made is likely to be a waste of time and money if general practice concepts are not taken into account more explicitly.

It is helpful to know that the main reason for starting the current standards policy of the Dutch College of General Practitioners was not a perceived need to increase quality in general practice, but more fear of governmental intervention (for budgetary reasons) in the autonomy of general practitioners. The College thought it was better to start a quality initiative itself than to be forced into accepting governmental supervision on this issue. Nevertheless this initiative may well lead to good results with regard to improving quality, if the standards are accompanied with quality assurance programmes. Standard setting as a target in itself will never succeed in raising the quality in general practice. Standard setting committees should be aware of the phenomenon of "arm-chair standards", a term used to describe the danger that committees, in their desire to produce consensus standards, design standards which do not have much to do with day-to-day practice.[19]

Our results show that although doctors did not perform according to the standards in the performance part, they were able to act substantially better in this respect in the competence part, but still not fully at the level of the standards. This finding implies at least three hypotheses.

Firstly, the doctors are not able to give good quality of care, since they fail to reach the standards even in the competence part.

Secondly, the standards were not realistic, meaning that they did not take account of the actual capacities of doctors in the performance part as well as in the competence part.

The third hypothesis is a mixture of the first two: doctors do not perform well enough and the standards will have to take more account of day-to-day practice. We tend to think that the third hypothesis will be the one most likely to be true.

We therefore think that more research is necessary with regard to standards in general practice. We would like to suggest conducting more studies to find out what happens in actual consultations in general practice. This can be done either by standardized patients or by other methods which are able to clarify the real practice of doctors, such as videotapes of consultations.[20] We would like to suggest emphatically that this kind of research will also have to be focussed on the reasons why doctors perform as they do, for example by asking doctors about the reasons they have for carrying out or not carrying out certain actions during a consultation. Another way to find out whether doctors have good reasons to perform as they do in real practice is to compare the findings on the content of consultations with findings in the literature about the medical content of such consultations. It is sad that most of the current research (at least in the Netherlands) with regard to standards

is based mainly on questionnaires to doctors, asking whether they support the standards programme and whether they are aware of the content of standards. This type of research has serious shortcomings, as we have showed in this study.[21]
As long as research with regards to standards is still in its infancy, we would like to state that when it comes to the assessment of actual care for licensure reasons, current standards cannot yet be used.

More research should also be addressed to the hypothesis that the implementation of standards in actual practice indeed leads to a better outcome for patients, because in the end the only good reason for working according to standards for general practitioners will have to be that it produces better outcomes.

## References


1. Senior JR. Towards the measurement of competence of medicine. Philadelphia: National Board of Medical Examiners 1976.
2. Lloyd JS. Definitions of competence in specialties of medicine. Chicago: American Board of Medical specialties 1979.
3. Viz. Stillman PL. Personal communication.
4. Cate ten RS. Simulatie 1. en Verbrugh HS. Simulatie 2. Letters to the Editor. Huisarts Wet 1988; 31: 141-2.
5. Rethans JJE. Standards for primary care. (Letter to the editor). J Fam Prac 1988; 26: 248-9.
6. Vossen JJJH. 'Modern Times'. Letter to the Editor. Med Contact 1991; 46: 133-4.
7. Watts G. Medicine Now. Standardized patients in real practice. British Broadcast Corperation 4; March 12 1991: 11.00-11.10 am.
8. Rethans JJ. Validity of case simulations. Letter to the Editor. J Clin Epidemiol; in press.
9. Stillman PL. Ruggil JS, Rutala PJ, Sabers DL. Patient instructors as teachers and evaluators. J Med Educ 1980; 55: 186-93.
10. Stillman P, Ruggill J, Rutala P,Sabers D. An instructional program using patient instructors as teachers and evaluators. Proc Annu Conf Res Med Educ 1979; 18: 53-8.
11. Stillman PL, Ruggill JS, Rutala PJ, Sabers DL. Patient instructors as teachers and evaluators. J Med Educ 1980; 55: 186-93.
12. Schmidt HG, Norman GR, Boshuijzen HPA. A cognitive perspective on medical expertise: Theory and implications. Acad Med 1990; 65 : 611-21.
13. Schmidt HG, Boshuijzen HPA, Hobus PPM. Transitory stages in the development of medical expertise: the intermediate effect in clinical case representation studies. In Proc. 10th Conf Cognitive Science Soc. Hillsdale, New Jersey: Erlbaum; 1988: 139-45.
14. Norman G, Allery l, Berkson L et al. Research in the psychology of clinical reasoning: implications for assessment. In: Jolly B (ed). New directions in the assessment of clinical competence. Proc Cambridge Conf IV, Madingley Hall, Cambridge. In press.
15. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision making in pulmonary nodule detection. Invest. Radiol. 1987; 13: 175-81.
16. Hobus PPM, Schmidt HG, Boshuizen HPA, Patel VC. Contextual factors in the activation of first hypotheses: expert-novice differences. Med Educ 1987; 21: 471-76.
17. Vleuten van der CPM, Verwijnen GM. A system for student assessment. In: Vleuten van der C, Wijnen van W (eds). Problem-based learning. Perspectives from the Maastricht Experience. Amsterdam 1990; Thesis Publishers.
18. Hobus PPM, Hofstra ML, Boshuizen HPA, Schmidt HG. De context van de klacht als diagnosticum. Huisarts Wet 1988; 31: 261-67.

19. Neufeld VR , Norman GR. Assessing clinical competence. Springer Publishing Company, New York, 1985.
20. Pringle M, Stewart-Evans C. Does the awareness of being video recorded affect doctors' consultation behaviour? Br J Gen Pract 1990; 40: 455-8.
21. Rethans JJE, Van Boven CPA. Simulated patients in general practice: a different look at the consultation. Br Med J 1987; 294: 809-12.

## Summary

The introduction of this book **(chapter 1)** provides the reader with the reasons why the studies, described here, have been conducted.

The immediate instigation was the finding in literature that there might be a discrepancy between what doctors do in their practice and what they tell you they do in an interview or in a questionnaire. It is also described why there were reasons to hypothesize that doctors can do better in a test situation than in actual practice. These questions called for a method, which was able to establish the performance in real practice of general practitioners. It is described why the method of standardized patients was chosen for our studies. Reasons are given for the definitions used in this book, i.e. performance is "what a doctors does in his day-to-day practice" and competence is "what a doctor is capable of doing".

Two main purposes are described. The first was to investigate whether it was possible for a set of standardized (or simulated) patients to visit general practitioners during normal surgery hours, without being detected. The second purpose was to investigate 1) the difference and 2) the relationship between competence and performance of general practitioners.

**Chapter 2** presents the results of the first experiment with standardized patients in actual practice. Standardized patients made appointments with 48 general practitioners during actual surgery hours and collected facts about their performances. The standardized patients were indistinguishable from real patients and presented a standardized story of a symptomatic urinary tract infection. Two months later the same general practitioners received a written simulation about a patient who had the same urinary tract infection and were asked how they would handle this problemin real practice. Both results were scored against an existing consensus standard.

The overall score for both methods did not show any substantial differences. A more differentiated analysis, however, showed that general practitioners performed significantly better with standardized patients. It is also showed that general practitioners answering the written simulation performed significantly more superfluous actions. The results of this experiment show that the use of standardized patients seems to show the efficient performance of general practitioners.

**Chapter 3** presents the results of the same experiment as was described in chapter 2, but now focussing on the feasibility of sending standardized patients into doctors offices. This chapter also provides the reader with the actions, as they were performed by the general practitioners in real practice. These data were scored according to a consensus standard. It is concluded that the method of introducing standardized patients into doctors offices was

feasible and that the participating physicians performed only 60% of the actions, which were defined to be essential in the standard used. These results are discussed.

**Chapter 4** presents a review on the use of standardized patients. Main purpose of this chapter was to search for the purposes for which standardized patients have been used. Also the feasibility, reliability and validity of the use of standardized patients was investigated.

It is concluded that for educational purposes standardized patients are a unique method and that the validity of the method is good. Concerning the reliability of the use of standardized patients for educational purposes, it is concluded that research on reliability has only recently started and that the results are promising, but more research has to be conducted. The use of standardized patients for practice purposes shows promising results on reliability and validity assesment.

**Chapter 5**, a review paper based on the 18 most important studies in the literature about medical competence, tries to re-initiate the debate about "what is a competent general practitioner?" by proposing clear distinction between 'competence', (what a physician is capable of doing) and 'performance' (what a physician does in his day-to-day practice). With this distinction we looked at whether studies defined both competence and performance, how they dealt with these concepts, what measurement instruments were used and what the conclusion of the studies were.

Although it is the common reasoning that competence is a good predictor of performance, this concept could not be affirmed. The survey showed that the majority of studies use wrong concepts and come to invalid conclusions. With the empirical distinction between competence and performance however, this chapter proposed new directions for quality assessment of general practitioners.

**Chapter 6** describes in detail the method to introduce standardized patients into general practice consultations. The chapter looked first at the general preparation by which the method was organized regardless of individual practices or doctors and secondly at the specific preparation concerning the fine detail of the individual visit. The method was tested in 156 consultations with 39 general practitioners and in no cases were the standardized patients detected. It was concluded that the standardized patient method, following the step-by-step procedure described, is feasible in actual practice.

**Chapter 7** represents the results of a study in which 39 general practitioners were each visited during normal surgery hours by four standardized patients. The objective of the study was to see whether actual

performance of general practitioners, as assessed by standardized patients, met predetermined consensus standards of care for actual practice.

The mean group scores of the doctors on the standards ranged from 33 to 68%. The results show that standardized patients may be the method of choice in the assessment of the quality of actual care of doctors. It is hypothesized that the substandard scores of the doctors do not reflect inadequate competence, but are a result of the difference between competence and performance.

**Chapter 8** shows the results of a study, which purposes it was to investigate 1) the difference and 2) the relationship between performance and competence. 39 Family physicians were consulted by four, incognito, standardized patients, portraying four different cases, during normal surgery hours. Later 34 of the 39 doctors participated in a controlled practice test at the Medical School, for which they were asked to act to the best of their ability. Here they were confronted with exactly the same standardized cases (different patients) as they had seen earlier in real practice. Based on the reports of the standardized patients, scores were assigned to the physicians.

Results showed that mean competence score of the physicians was substantially higher than their mean performance score. The Pearson correlation between the competence score and the performance score of the participating physicians was negligable. However, when efficiency and consultation time of the consultations were taken into account, moderate to high correlations were found. It was concluded that performance and competence should be considered as distinct constructs. Assessment of competence under examination circumstances can only have predictive value for performance in actual practice when factors such as efficiency and consultation time are taken into account. Below standard performance of physicians does not necessarily reflect a lack of competence.

**Chapter 9** rehearses the main conclusions of the research described in this book.

With respect to the method of standardized patients in real practice, it is concluded that this method has proved to be feasible and reliable. For further research with standardized patients in real practice it would be worthwhile to extend the reports of standardized patients also to the social and attitudinal skills of doctors. It is also adviced to use the method for the study of interdoctor variation.

Concerning the relationship between competence and performance it is concluded that physicians are able to perform more actions in a competence setting compared with the performance setting. On the other hand if efficiency is taken into account, physicians do better in real practice than in a test setting. It is concluded that competence and performance should be considerd as two distinct constructs.

With regard to the issue whether competence is a predictor of performance, it appears to be crucial to choose the correct variable to measure the correlation between competence and performance. It is concluded that the factor 'time' seems to play a major role in this. It is adviced that more research is necessary to be able to estimate the importance of this time factor.

Concerning standards of care it is concluded that the participating physicians performed below standards in the performance part. On the other hand, they performed considerably better in the competence part, but still below standards of care. More research is necessary to study the validity of standards. At present standards are not (yet) an instrument to assess the quality of performance of physicians in real practice. For educational purposes, however, they are good material.

## Samenvatting

De introductie van dit boek **(hoofdstuk 1)** bevat de redenen voor het beschreven onderzoek. Aanleiding tot dit onderzoek was een literatuur-referentie, die de hypothese bevatte dat er een verschil bestaat tussen wat artsen in hun dagelijkse praktijk doen, en wat ze in interviews of in antwoord op een vragenlijst zeggen te doen. Daarnaast geeft de introductie de achtergronden van de hypothese dat artsen in een testsituatie beter kunnen handelen dan in de dagelijkse praktijk. Beide hypothesen vereisten een onderzoeksmethode die het feitelijk handelen (performance) van artsen kon vastleggen. Hier wordt beschreven waarom de simulatiepatiënt-methode hiervoor gekozen is.

Tevens bevat de introductie een verantwoording van de gehanteerde definities van de begrippen 'performance' en 'competence' (resp. 'dat wat een arts doet in zijn dagelijks praktijk-handelen' en 'dat waar een arts toe in staat is').

Ten slotte worden de twee hoofddoelen van het onderzoek aangegeven. Het eerste was te onderzoeken of het voor simulatiepatiënten mogelijk was, om huisartsen tijdens hun dagelijkse spreekuren onontdekt te bezoeken. Het tweede doel was, 1) het verschil, en 2) de relatie tussen competence en performance van huisartsen te onderzoeken.

**Hoofdstuk 2** vermeldt de resultaten van het eerste experiment met simulatiepatiënten in de feitelijke praktijk. Deze simulatiepatiënten maakten afspraken bij 48 huisartsen, en verzamelden gegevens over deze echte consulten. Deze patiënten waren getraind in het spelen van een gestandaardiseerde rol: die van een vrouw met een symptomatische urineweginfectie. Twee maanden na deze bezoeken ontvingen de bezochte artsen een schriftelijke simulatiepatiënt-casus met dezelfde inhoud als tijdens de eerdere praktijkbezoeken. Daarbij kregen zij de vraag voorgelegd hoe ze met een dergelijke patiënt in de praktijk zouden omgaan. Hun handelen werd gescoord volgens een consensus-standaard.

Een globale analyse lijkt geen verschil tussen deze twee methoden aan te geven. Een gedifferentieerde analyse geeft echter aan dat de deelnemende artsen het tijdens het echte consult significant beter deden dan op schrift. Op het punt van niet-noodzakelijke en overbodige handelingen verrichtten de artsen meer handelingen op schrift dan in werkelijkheid. De resultaten geven aanleiding, een efficiënt spreekuurgedrag van artsen te veronderstellen.

**Hoofdstuk 3** geeft andere resultaten weer van het in hoofdstuk 2 besproken onderzoek, - nu met de nadruk zowel op de uitvoerbaarheid van het gebruiken van simulatiepatiënten voor praktijkonderzoek, als op de in de praktijk verzamelde gegevens. Dit leidt tot de conclusie dat de

simulatiepatiënt-methode bruikbaar lijkt om data tijdens spreekuren te verzamelen. Tevens blijkt dat de huisartsen slechts 60% van de in de standaard als noodzakelijk gedefinieerde handelingen verrichtten. Deze resultaten worden bediscussieerd.

**Hoofdstuk 4** is een literatuuronderzoek naar de doeleinden waarvoor simulatiepatiënten door de jaren heen zijn gebruikt, en naar de betrouwbaarheid en de validiteit van dit gebruik. De conclusie luidt dat de simulatiepatiënt-methode voor educatieve doeleinden een uniek leermiddel is gebleken en dat de validiteit goed is. Betreffende de betrouwbaarheid van simulatiepatiënten voor educatieve doeleinden valt te constateren dat de betreffende resultaten veelbelovend zijn, maar dat hiernaar meer onderzoek nodig is. Het gebruik van simulatiepatiënten voor praktijkdoeleinden is op dit punt beter onderzocht en de resultaten zijn goed.

**Hoofdstuk 5** is een literatuuronderzoek, gebaseerd op de achttien belangrijkste studies naar medische competentie. Hiervoor is onderscheid gemaakt tussen 'performance' (dat wat een arts doet in zijn dagelijks praktijk-handelen) en 'competence' (dat waartoe een arts in staat is). Met dit onderscheid zijn de achttien studies bekeken op de vragen : of, en zo ja, hoe zij deze begrippen definiëren; welke meetinstrumenten gebruikt zijn; en wat de conclusies van de studies zijn.

Hoewel algemeen de gedachte heerst dat competence een goede voorspeller van performance is, wordt deze gedachte niet bevestigd. Het onderzoek toont aan dat de meerderheid van de onderzochte studies de verkeerde begrippen gebruikt en niet-valide conclusies trekt. Ter afsluiting geeft het hoofdstuk nieuwe wegen aan om de kwaliteit van huisartsgeneeskundig handelen vast te stellen.

**Hoofstuk 6** beschrijft gedetailleerd de stappen die nodig zijn om simulatiepatiënten een bezoek te laten brengen aan huisartsen in de echte praktijk. Eerst wordt de algemene voorbereiding beschreven, los van kenmerken van individuele praktijken. Dan volgt een beschrijving van de specifieke voorbereiding, die van praktijk tot praktijk verschilt.

Deze methodiek is getest in 156 consulten bij 39 huisartsen, waarbij geen enkele simulatiepatiënt ontdekt is. De conclusie is dat de simulatiepatiënt-methode voor gebruik tijdens echte spreekuren, volgens de beschreven gefaseerde aanpak, zeer geschikt is.

**Hoofdstuk 7** geeft de resultaten weer van een onderzoek waarin 39 huisartsen, tijdens hun spreekuren ieder door vier verschillende simulatiepatiënten werden bezocht. Het doel hiervan was, te onderzoeken hoe artsen in hun dagelijkse praktijk handelen ten opzichte van vastgestelde consensus-standaarden.

Het blijkt dat artsen per standaard gemiddeld 33 tot 68% van de handelingen verrichten die volgens de standaarden obligaat zijn. Dit leidt tot de hypothese dat het niet volgens standaarden werken geen gevolg is van onvoldoende competentie, maar eerder het gevolg is van een verschil tussen competence en performance.

**Hoofdstuk 8** bespreekt de resultaten van een onderzoek dat tot doel had, 1) het verschil en 2) de relatie tussen competence en performance van huisartsen te onderzoeken. Hiertoe werden 39 huisartsen tijdens hun spreekuren bezocht door vier simulatiepatiënten, die elk een andere klacht naar voren brachten. Deze simulatiepatiënten waren als zodanig niet te herkennen. Hierna deden 34 van de 39 huisartsen mee aan een gecontroleerde test op het skillslab van de universiteit, waarbij de artsen het specifieke verzoek kregen, kwalitatief zo hoog mogelijk te handelen. Tijdens deze test werden de artsen geconfronteerd met simulatiepatiënten met dezelfde casuïstiek als de artsen eerder in de praktijk was aangeboden. Op grond van de rapportages van de simulatiepatiënten zijn per arts scores berekend.

De resultaten geven aan dat de gemiddelde competence-score van de artsen significant hoger is dan de gemiddelde performance-score. De Pearson-correlatie tussen de competence- en de performance-score is verwaarloosbaar laag. Doch bij verdiscontering van de efficiëntie en de consultduur in de scores zijn goede tot hoge correlaties te vinden.

De conclusie luidt dat competence en performance als twee verschillende begrippen moeten worden beschouwd. Het meten van competentie onder examen-omstandigheden heeft alleen voorspellende waarde voor de dagelijkse praktijk, als efficiëntie en consultduur ook medebepaald worden. Het niet volgens standaarden handelen hoeft geen gebrek aan competentie te betekenen.

**Hoofdstuk 9** herhaalt de belangrijkste conclusies van het onderzoek.

Wat de simulatiepatiënt-methode voor gebruik in de dagelijkse praktijk betreft, luidt de conclusie dat deze methode bruikbaar, betrouwbaar en valide is. Aanbevolen wordt, meer onderzoek te doen met simulatiepatiënten in de dagelijkse dokterspraktijk, zowel naar het rapporteren van sociale vaardigheden en attitude-vaardigheden, als naar interdokter-variatie.

Betreffende de verhouding tussen competence en performance is de conclusie dat artsen meer handelingen verrichten in een competence-situatie dan in een performance-situatie. Als echter ook naar efficiëntie wordt gekeken, doen artsen het beter in de praktijk dan in de testsituatie. De gevolgtrekking is dat competence en performance als twee verschillende begrippen moeten worden beschouwd.

Inzake de schatting van de correlatie tussen competentie en praktijk-handelen, blijkt dat het van wezenlijk belang is, hiervoor de juiste variabele

te kiezen. De conclusie op dit punt is dat de consultduur een belangrijke rol speelt en dat meer onderzoek hiernaar nodig is.
Tevens is geconcludeerd dat artsen in hun dagelijks praktijk-handelen niet volgens standaarden handelen. In een testsituatie doen ze het wel beter, maar nog steeds niet volgens de standaarden. Er is meer onderzoek naar de validiteit van standaarden nodig, voordat deze bruikbaar zijn als maat voor de kwaliteit van handelen. Voor onderwijskundige doeleinden zijn standaarden echter wel een goed leermiddel.

## Resymé

Innledningen i denne boken (**kapittel 1**) behandler bakgrunnen for den beskrevne undersøkelsen.

Foranledningen til denne undersøkelsen var en litteraturhenvisning, som inneholdt den hypotese at det er en forskjell på hva leger foretar seg i sin daglige prakis og hva de sier i intervjuer eller i sine svar på spørreskjemaer, at de foretar seg. Dessuten gir innledningen bakgrunnene for den hypotese at leger i en forsøkssituasjon kan handle bedre enn i den daglige praksis. Begge hypoteser krevde en undersøkelsesmetode som kunne registrere legers faktiske fremgangsmåte ('performance'). Her beskrives hvorfor en undersøkelsesmetode ved bruk av simulerte pasienter er valgt til dette formålet.

Videre inneholder innledningen en redegjørelse for de brukte definisjoner av begrepene 'performance' og 'competence' (hhv. 'det som en lege gjør i sin daglige yrkesmessige virksomhet' og 'det som en lege er i stand til å gjøre').

Til sist nevnes undersøkelsens to hovedformål. Det første var å undersøke om det var mulig for simulerte pasienter å besøke praktiserende leger i deres daglige konsultasjonstid uten å bli oppdaget. Det andre formål var å undersøke 1) forskjellen, og 2) forholdet, mellom praktiserende legers 'competence' og 'performance'.

**Kapittel 2** nevner resultatene av det første eksperimentet med simulerte pasienter i den faktiske praksisen. Disse simulerte pasientene fikk avtaler hos 48 praktiserende leger, og samlet inn data om disse faktiske konsultasjonene. Disse pasientene hadde blitt lært opp til å spille en standardisert rolle: som en kvinne med en symptomatisk urinveisinfeksjon. To måneder etter disse besøkene mottok de konsulterte leger et skriftlig kasus med det samme innholdet som under de tidligere faktiske konsultasjonene med de simulerte pasientene. I denne forbindelse ble de spurt om å beskrive hvordan de ville behandle en slik pasient i praksis. Deres handlemåte ble målt etter en konsensus-standard.

En grov analyse synes ikke å vise noen forskjell på disse to metodene. En differensiert analyse viser imidlertid at de deltakende legene handlet signifikant bedre under den virkelige konsultasjonen enn i sine skriflige svar. Med hensyn til unødvendige og overflødige handlinger skrev legene at de utførte flere av disse enn det viste seg at de gjorde i virkeligheten.

Resultatene gir grunn til å anta en effektiv handlemåte i konsultasjonstiden.

**Kapittel 3** gjengir andre resultater av undersøkelsen diskutert i kapitel 2, - denne gangen både når det gjelder muligheten til å gjennomføre bruk av simulerte pasienter til prakis-undersøkelser, og når det gjelder de faktisk

innsamlede data. Dette fører til den konklusjon at metoden med simulerte pasienter synes å kunne brukes til innsamlingen av data i konsultasjonstiden. Desuten viser det seg at de praktiserende legene utfører bare 60% av de handlinger som i standarden defineres som nødvendige. Disse resultater diskuteres.

**Kapittel 4** er en litteraturundersøkelse av de formålene simulerte pasienter har vært benyttet til i årenes løp, og av påliteligheten og validiteten av å bruke simulerte pasienter.

Konklusjonen er at bruk av simulerte pasienter til undervisningsformål har vist seg å være et enestående læremiddel og at validiteten er god. Med hensyn til påliteligheten av simulerte pasienter til undervisningsformål kan det fastslås at de aktuelle resultatene er lovende, men at det er nødvendig med ytterligere forskning på dette område. Bruk av simulerte pasienter til praktiske formål er blitt bedre undersøkt på dette punkt, og resultatene er gode.

**Kapittel 5** er en litteraturundersøkelse, som er basert på de atten viktigste studier i lege-kompetanse.

I denne forbindelse skjelnes det mellom 'performance' (det som en lege gjør i sin daglige yrkesmessige virksomhet) og 'competence' (det som en lege er i stand til å gjøre).

På bakgrunn av denne forksjellen er de atten studier blitt undersøkt med hensyn til følgende spørsmål: om, og i så fall, hvordan, de definerer disse begrepene; hvilke måleinstrumentene som er blitt brukt; og hvilke konklusjoner studiene trekker. Skjønt det er en generell oppfatning at 'competence' gir en god forutsigelse av 'performance', bekreftes denne oppfatningen ikke. Undersøkelsen viser at de fleste studier bruker uriktige begreper og trekker ikke-valide konklusjoner.

Til slutt viser kapitlet nye veier til å fastslå kvaliteten av legers fremgangsmåte.

**Kapittel 6** gir en detaljert beskrivelse av de trinn som er nødvendige for å få simulerte pasienter til å konsultere praktiserende leger i virkelig praksis.

Først beskrives den alminnelige forberedelsen, uten hensyn til kjennetegn av individuelle legers praksis. Deretter følger en beskrivelse av den spesifikke forberedelsen, som er forskjellig fra praksis til praksis. Denne metodikken er blitt prøvd i 156 konsultasjoner hos 39 praktiserende leger, og ingen simulert pasient er blitt oppdaget.

Konklusjonen er at metoden med simulerte pasienter, til bruk under virkelige konsultasjoner, etter den beskrevne trinnvise metoden, egner seg meget godt.

**Kapittel 7** gjengir resultatene av en undersøkelse der 39 praktiserende leger hver ble besøkt av fire forskjellige simulerte pasienter i konsultasjonstiden.

Formålet med dette var å undersøke hvordan leger i sin daglige praksis handler sammenlignet med faste konsensus-standarder. Det viser seg at leger pr. standard utfører i gjennomsnit 33 til 68% av de handlingene som ifølge standardene er obligatoriske.

Dette fører til den hypotese at avvikelsen fra standarden ikke er en følge av utilstrekkelig kompetanse, men snarere en følge av en forskjell på 'competence' og 'performance'.

**Kapittel 8** diskuterer resultatene av en undersøkelse met det formål, å undersøke 1) forskjellen, og 2) forholdet, mellom praktiserende legers 'competence' og 'performance'.

I denne forbindelse ble 39 praktiserende leger besøkt av fire simulerte pasienter, som hver for seg kom med forskjellige symptomer. Disse simulerte pasientene var ikke til å gjenkjenne som sådan. Heretter deltok 34 av de 39 praktiserende leger i en kontrollert prøve i ferdighetslaboratoriet på universitetet, der de uttrykkelig ble spurt om å handle kvalitativt høyest mulig. Under denne prøven ble legene konfrontert med simulerte pasienter med de samme symptomene som legene tidligere var blitt konfrontert med i sin praksis. På grunn av de simulerte pasienters rapporter er det blitt beregnet poengsum pr. lege.

Resultatene viser at legenes 'competence'-score i gjennomsnitt er signifikant høyere en den gjennomsnitlige 'performance'-score. Pearson-korrelasjonen mellom competence'- og 'performance'-scoren er ubetydelig lav. Men dersom effektiviteten og konsultasjonsvarigheten i disse 'scores' tas i betraktning, finnes det fra gode til høye korrelasjoner.

Konklusjonen er at 'competence' og 'performance' må betraktes som to forskjellige begrep. Måling av kompetanse under eksamensomstendigheter har bare forutsigelsesverdi for den daglige praksis, når effektiviteten og konsultasjonsvarigheten også tas i betraktning. Det at legen ikke handler etter standarder behøver ikke å bety mangel på kompetanse.

**Kapittel 9** gjentar undersøkelsens viktigste konklusjoner.

Vedrørende undersøkelsesmetoden med simulerte pasienter til bruk i den daglige prakis, er konklusjonen at denne metoden er formålstjenlig, pålitelig og valid. Det anbefales å forske mer med simulerte pasienter i den daglige praksis, både hva angår rapportering av sosiale ferdigheter/holdningsferdigheter, og lege-variasjon.

Med hensyn til forholdet mellom 'competence' og 'performance' er konklusjonen at leger utfører flere handlinger i en 'competence'-situasjon enn i en 'performance'-situasjon. Tas imidlertid også effektiviteten i betraktning, handler leger bedre i praksis enn i forsøkssituasjonen.

Konklusjonen er at 'competence' og 'performance' må betraktes som to forskjeligge begrep.

Hva angår vurderingen av korrelasjonen mellom kompetanse og praktisk handling viser det seg å være av vesentlig betydning å velge den rette variabelen. Konklusjonen på dette punktet er at konsultasjonsvarigheter spiller en viktig rolle og at ytterligere forskning på dette området er nødvendig.

En anden konklusjon er at leger i sin daglige praksis ikke handler etter standarder. I en forsøkssituasjon gjør de det riktignok bedre, men likevel ikke etter standardene.

Ytterligere forskning av standarders validitet er nødvendig, før disse lar seg bruke til å måle handlingskvaliteten. Til undervisningsformål er standarder imidlertid et godt læremiddel.

## Thanks

Successful research depends on many people.
I have asked for and received the help and advice of a considerable number of persons. Many of them work in the field of research, others in the field of general practice. I am also most grateful for the assistance and support which I have received from quite a number of people outside the academic world, such as my family and friends.

This study would have been impossible without the participating general practitioners, whose names I have promised not to mention, and without the help and advice of Ferd Sturmans, Riet Drop, Cees van der Vleuten, Trudie Seegers and Geoff Norman (McMaster University, Hamilton, Canada).

Dear friends, I would like to thank you all for helping me during the stressful and exciting, but at the same time very pleasant years that are now behind me. I hope that it will be possible to continue to cooperate with many of you in the future.

I would like to mention (in alphabetical order):
Pierre Bastings, Els Boshuizen, Els Bour, Cees van Boven, Dhr. Braams, Pie Castermans, Michiel Cornel, Donald Crombie (Birmingham, UK), Pia Cuijpers, Dhr. van Deelen, Geert-Jan Dinant, Frans Duijsens, Yvonne L'Espoir-Dohmen, Toon van Gerven, Gale Glivia (McMaster University, Hamilton, Canada), Erik de Graaf, Diederik Grit, Richard Grol, Bea de Groote, Theo Hellemons, Leon Heuts, Pie Hobus, Ron Hoogenboom, Paul Höppener, Robin Hull (Birmingham, UK), Caroliene Janssens, Jan-Willem de Jonge, Victor Kaiser, Juul Kerbusch, Piet Kerkhof, Jan Klerkx, Andre Knottnerus, Peter Kramer, Marike Laning, Yvonne van Leeuwen, Jan Leewis, Martha Lucassen, Wil Macco, Jaap Metz, Jean Muris, Annemarie Muysken, Paul Nekeman, Frank van Onna, Charles Phaff, Marijke Perquin, Piet Portegijs, Pieter Ramler, Jan C.A. Rethans, Thea Rethans-Willems, Henk Schmidt, the sickfund SVGZ in Maastricht (Board of Directors and Dhr J. Schoffelen in particular), the sickfund LIASS in Heerlen and Sittard (Board of Directors and Dhr J. Quadvlieg and Dhr. Rohs in particular), Richard Starmans, Jelle Stoffers, Jacques van Thiel, Marijke Verdonk, Theo Voorn, Maarten Verwijnen, George Wolfs, Wim van Zutphen and Paul Zwietering.

## Curriculum vitae

Jan-Joost Rethans was born on January 11th, 1955 in Delft, the Netherlands. He finished his secondary education (grammar school, specializing in sciences) at St. Stanislas' College in Delft in 1973. From 1973 to 1974 he studied economics at the Erasmus University in Rotterdam and from 1974 to 1975 he was a student at the medical school of the University of Ghent, Belgium.
From 1975 to 1982 he studied medicine at the University of Limburg in Maastricht. He spent the year1980-1981 in Grindaheim, Norway, working in general and veterinarian practice in a rural area. He graduated from the University of Limburg in 1982, remaining at the same university to start training as a general practitioner. In the period from 1983 to 1984 he was a general practitioner in several practices in the Netherlands. In 1984 he worked at the psychiatric hospital in Reinsvoll, Norway.
From 1985 to 1986 he participated the one-year Research Training Course for general practitioners at the University of Limburg.
Since 1986 he has been working as a researcher/general practitioner at the Department of General Practice of the University of Limburg, where he carried out the studies presented here. To learn more about standardized patients he spent five weeks (in 1986) at the Department of Clinical Epidemiology & Biostatistics of McMaster University in Hamilton (Canada), where he worked under the supervision of Prof. G.R. Norman.
Since 1987 he has been a member of the Executive Board of the European General Practice Research Workshop (EGPRW).
He is married with four children.

## Appendices

1. Detection form for the physicians to report standardized patients.
2. Instructions for the participating general practitioners (competence setting)
3. The standards of care used

## Detection form to report standardized patients

Date:
Name physician:
Address:
Postal Code:
City:

On what date did you detect the standardized patient?

What name did the standardized patient use to present him/herself?

Was the standardized patient a male/female?

For what main complaint did the standardized patients ask your help?

What reasons made you think that this was a standardized patient?

Did you detect the standardized patient before/during/after the consultation?

Do you believe that the detection of the standardized patient had a serious impact on your performance during that consultation?

Would you please encircle how certain you are that the patient, whom you are now reporting to us, is indeed a standardized patient?

Absolutely certain/ Certain/ Uncertain/ Absolutely uncertain

Further remarks?

Thank you for your cooperation.

## Instructions for participating general practitioners (Competence setting)

Dear colleague,

During the next two hours you will be confronted with one or more standardized patients. You have seen none of these standardized patients before. The problems for which the standardized patients are asking your help differ from the problems presented to you by the standardized patients during their visits in your actual practice. (If the problems were the same ones, you would now recognize them.)

### What is the purpose of this setting?

It has often been assumed that general practitioners deliver sub-optimal patient care in their practice because they do not know how to do better. The purpose of this study is to investigate if that is true. We believe that general practitioners are quite capable of treating their patients well, but that normal practice circumstances (busy surgery hours, telephone calls during contacts, etc) may prevent them providing optimal care.

### Top quality

We would like you to consider the room in which you are now sitting as your own surgery. We would also like you to treat the standardized patient or patients who will be coming to you in a moment with the **highest possible degree of quality** (anno September 1989). Show the best you are capable of!

We want you to perform to the best of your abilities during this surgery hour. It may be that this will mean that you will be performing better than in your routine surgery hours. We would like to assess the quality of general practitioners' performance when they are working at their top level. This top level performance is required during all parts of a consultation, not only for the history and physical examination but also for the choice of prescription and for concluding the contact.

In this surgery hour you will be free of time pressure. We will assess only the quality of the content of your performance and will not pay attention to the time used. There is no difference for us whether you perform at the same top level of quality in 5 minutes or in 25 minutes. We know from previous experience that much longer than 25 minutes will probably not be neccessary.

**Practical notes.**
From now on, you are a general practitioner in this room. It is morning. If you are used to working in a health centre, then please consider this room as a room in your health centre. If you are used to working in a solo practice, then you are also working in a solo practice now.

**Chart sytem**
Each of the patients to come represents a new patient to you. Your medical secretary has already listed this patient in your practice and she has already made up a patient file (medical chart). This will be handed to you by the patient him/herself. You do not have to bother about details such as the visiting standardized patient's previous general practitioner.
We would like to ask you to fill in on this medical chart (during or after the consultation) those items which you consider absolutely necessary. This medical chart should also reflect the top performance of this consulation.

**Laboratory examination**
We would like you to complete each consultation. If you wish the patient to go to a hospital laboratory (for a blood test, for example), then please fill in a lab form. The lab forms are in front of you and are the same as the ones you are used to. After handing the form to the patient, you will end the consultation, again in the top-level manner. If, on the other hand, you are used to doing lab tests in your own laboratory, then you may ask the medical student, who will be sitting in this room too, for the result of the test. The result will than be given to you. For example, if you would like to assess the hemoglobin level of the patient or to do a urine test, you may ask the student: "I want to have the hemoglobine value, what is it?" or "I would like to check something in the patient's urine, what are the results?". After having received this information you may continue with the consultation. If you would like to prescribe a drug or drugs to the standardized patient, please write out a prescription and hand it to the patient.

**Medical instruments**
We have tried to provide this room with all the usual medical instruments, for instance a stethoscope, a Mercury sphygmomanometer, a 'Snellen' chart, etc. Please take a good look around the room to get acquainted with it. Also take a good look at the instruments. If you think that there is something missing and if you would like to use this device for the examination, then please pretend you are using it and say so clearly to the patient.

In a few moments the first patient will knock on your door. When you have finished the contact, the patient will leave the room. We would like you not to start discussing the contact with the standardized patient afterwards, since he/she has to fill in a form immediately after the contact.

It is may well be that you will have to wait some time before the next patient enters. Please stay in this room. We have supplied you with a newspaper. If you have urgent questions, please ask the student and he/she will ask me in the corridor.
Once again I would like to thank you for your cooperation in this study.

**Good luck!**

**Diagram 1. Standard of the Headache case**

**History**

1. Obligatory. Time aspects: how long have complaints existed; at what times of the day; how often?
2. Obligatory. Nature of the pain.
3. Obligatory. Location of the pain, radiation, presence of prodomes; progression of complaints.
4. Obligatory. Associated phenomena (such as light phobia, nausea, fever, dental and neck complaints,etc.).
5. Obligatory. Relation with psychosocial circumstances.
6. Intermediate. What is the reaction to the pain.
7. Intermediate. Self therapy.
8. Intermediate. Intoxications (smoking, carbon monoxide).
9. Obligatory. Migraine history (familial history, progression, frequency, connection with circumstances).
10. Superfluous actions history.

**Physical examination**

11. Intermediate. Blood pressure reading.
12. Intermediate. Examine eye sight.
13. Intermediate. Examine eye fundus.
14. Intermediate. Cervical spine.
15. Intermediate. Neurological examination.
16. Intermediate. Sinuses: percussion and pressure pain, transillumination.
17. Superfluous actions physical examination.

18. Superfluous actions Laboratory

**Guidance and advice**

19. Obligatory. Explain cause of complaints.
20. Obligatory. Discuss prognosis.
21. Obligatory. Explain relationship between complaints and tension.
22. A. Obligatory. In case of therapy: explain expected effect.
    B. Obligatory. In case of no therapy: explain why no therapy is prescribed.
23. Intermediate. Relaxation exercises, yoga (brochure).
24. Obligatory. Discuss connection with life style.
25. Superfluous actions guidance and advice.

**Therapy**

26. Intermediate. Simple analgetics.
27. Obligatory. Discuss possible background to headache.
28. Intermediate. Benzodiazepines.

29. Superfluous actions therapy.

**Return visit**

30. Obligatory. Indicate whether or not a return visit is necessary, depending on possible increase of complaints and prognosis.
31. Superfluous actions return visit.

**Diagram 2. The standard of the diarrhea case**

**History**

1. Obligatory. Presence of diarrhea.
2. Obligatory. Presence of nausea, vomiting.
3. Obligatory. Presence of abdominal pain.
4. Obligatory. Course and duration of the complaints.
5. Obligatory. Ask for possible causes such as: contact with illness; travel; special foods, etc.
6. Obligatory. Are there any other complaints?
7. Intermediate. Check medication.
8. Intermediate. Eating habits history.
9. Intermediate. Fever.
10. Obligatory. Stool consistency and frequency.
11. Obligatory. Nature of pain.
12. Obligaat. Location of pain, shifts in location.
13. Superfluous actions history.

**Physical examination**

14. Obligatory. Examine abdomen: inspection, percussion, auscultation, palpation.
15. Superfluous actions physical history.

**Other laboratory**

16. Obligatory. Stool culture.
17. Superfluous actions other laboratory.

**Guidance and advice**

18. Obligatory. Present diagnosis.
19. Obligatory. Discuss prognosis.
20. Obligatory. Give diet advice.
21. Superfluous actions guidance and advice.

**Medication**

22. Intermediate. Anti-emetics.
23. Intermediate. Adstringents.
24. Superfluous actions medication.

**Return visit**

25. Obligatory. Explain that patient should return in two or three days if complaints persist.
26. Superfluous actions return visit.

**Diagram 3. The standard for the Shoulder Pain case**

**History**

1. Obligatory. How long has the complaint existed?
2. Obligatory. Did the pain develop gradually or acutely?
3. Obligatory. Ask location of pain.
4. Obligatory. Can causative moments be found?
5. Obligatory. Is there any movement limitation; when, which movements?
6. Obligatory. How do complaints relate to life and work of patient?
7. Obligatory. Are there any accompanying symptoms such as paresthesias?
8. Intermediate. Ask about self-medication.
9. Intermediate. Are there any general illness symptoms?
10. Superfluous actions history.

**Physical examination**

11. Obligatory. Undress above waist, inspection.
12. Obligatory. Ask for pain locations to be indicated.
13. Obligatory. Palpation of pain points.
14. Obligatory. Palpate joint during movement.
15. Obligatory. Check passive movement: extension, flexion, rotation.
16. Obligatory. Check active movement: painful arc, flexed biceps, lower arm extensors.
17. Superfluous actions physical examination.

**Guidance and advice**

18. Obligatory. Report findings and or any diagnosis.
19. Obligatory. Discuss prognosis.
20. Obligatory. Give movement and exercise advice.
21. Intermediate. Give ergonomic advice.
22. Superfluous actions guidance and advice

**Medication and therapy**

23. Intermediate. Sling, immobilisation bandage, splint.
24. Intermediate. Antiphlogistics (NSAID); local corticosteroid injections.
25. Superfluous actions medication and therapy.

**Referral**

26. Intermediate. Physiotherapist.
27. Superfluous actions referral.

**Return visit**

28. Obligatory. Indicate whether or not a repeat visit is necessary, in relation with prognosis.
29. Obligatory. Return visit after 2-3 weeks.
30. Obligatory. Movement sufficiently limited to make working impossible.
31. Superfluous actions return visit.

**Diagram 4. The standard for the diabetes case**

**History**

1. Obligatory. Ask about health, hypoglycemias.
2. Obligatory. Ask about diet and medication problems.
3. Obligatory. Ask for cause of positive urine reduction >2%.
4. Obligatory. Pay attention once a year to sight complaints, neuropathy complaints, intermittent claudication complaints.
5. Superfluous actions history.

**Physical examination**

6. Obligatory. Determine weight.
7. Obligatory. Take blood pressure.
8. Obligatory. Fundus examination.
9. Obligatory. Inspect feet.
10. Obligatory. Palpation of arteries in feet.
11. Obligatory. Vibration sensee.
12. Superfluous actions physical examination.

**Own laboratory**

13. Obligatory. Pre-prandial blood glucose or at fixed time of day.
14. Obligatory. Morning urine protein level.
15. Intermediate. Urine glucose (stick).
16. Intermediate. Basic ECG.
17. Intermediate. Dipslide urine.
18. Intermediate. Determine ketones in urine.
19. Superfluous actions own laboratory.

**Other laboratory**

20. Obligatory. Cholesterol and serum lipids.
21. Intermediate. HbA1c.
22. Superfluous actions other laboratory.

**Guidance and advice**

23. Obligatory. Discuss suggested treatment.
24. Obligatory. Discuss diet.
25. Obligatory. Discuss cause of urine glucose.
26. Obligatory. Point out consequences of overweight.
27. Obligatory. Recommend weight loss.
28. Superfluous actions guidance and advice.

**Medication**

29. Obligatory. Discuss treatment policy.
30. Superfluous actions medication.

**Referral**

31. Intermediate. Refer to dietician.
32. Superfluous actions referral.

**Return visit**

33. Obligatory. Always schedule new appointment.
34. Obligatory. Return visit after 3-6 months.
35. Superfluous actions return visit.